Psychophysical models of masking for coding applications

Jont B. Allen
Room E161
AT&T Labs-Research
180 Park AV
Florham Park NJ 07932
973/360-8545voice, x8092fax
http://www.research.att.com/info/jba

February 1, 2005

Abstract

In this article we apply signal detection theory to qualitatively unify the *intensity just-noticable difference* (JND) and masking data of Wegel and Lane (1924), Fletcher and Munson (1933 – 1938), Miller (1947), and Egan and Hake (1950). We do this by treating the loudness as the first moment, and the intensity JND as the second moment, of the random variable we call the *single-trial loudness*. From these definitions we define a loudness signal-to-noise ratio ($SNR_{\mathcal{L}}$), which is the unifying factor. This theory relies heavily on Fletcher and Munson's 1933 theory of loudness. The purpose of this article is to create a model of masking that may be used for speech and music coding applications.

1 Introduction

1.1 The problem of perceptual coding

When quantizing signals one necessarily, by design, introduces noise into the representation. The art of perceptual coding is to control this noise in such a manner that has the smallest perceptual impact. Given a complete description of the signal dependent internal noise of the auditory system, it is assumed that it is possible to quantize the signals with a similar noise. Given such a coder, the quantizing error would be perceptually *masked* by the internal noise of the auditory system.

When the quantizing error is large enough that the error is above the perceptual threshold, we would like the system to degrade gracefully. How can we meet these difficult goals? The short answer is: only with a model of masking.

What is masking and where does it come from? How do we measure it experimentally? How may we predict it for an arbitrary signal? To understand the answers to these questions, we need models of loudness, masking, the intensity just-noticeable difference (JND_I), critical bands, and the listening condition. This article is about modeling masking and is an attempt to describe masking to the designer of a coder. It is not about how to engineer a speech or music coder.

When dealing with human perceptions we must carefully distinguish the *external* physical variables, which we call Φ variables, from the *internal* psychophysical variables, which we call Ψ variables (Note that Φ and Ψ sound like the initial syllable of the words *phy*sical and *psy*chological, respectively). The model we seek is a transformation from the Φ-domain to the Ψ-domain. Examples of Φ variables are pressure, frequency, intensity, and neural rate. Neural rate is an important example because it is an internal variable, yet it is physical; thus the terms *internal* and Ψ are not synonymous. Examples of Ψ variables are loudness, pitch, heaviness, brightness, tightness, and timbre.

To represent an acoustic signal, many Φ variables (e.g., intensity, frequency, duration, modulation) must be specified. In a similar manner, to understand and model masking, many internal Ψ variables must be determined. There is a natural but unfortunate tendency (everybody does it, all the time), to confuse the external physical (Φ) variables with the internal psychophysical (Ψ) variables (for example: heaviness and weight, pitch and frequency). Such confusions, in a scientific context, lead to serious misinterpretations of experimental results.

While the physical variables may be either deterministic or stochastic, it is essential to treat Ψ variables as stochastic. In fact, it is the stochastic character of the Ψ variables that is responsible for masking. Furthermore, to understand masking completely, it is important to describe it in both domains. Given these models, the Ψ-domain noise may be reflected back to the external Φ-domain, into the acoustic domain of the speech or music signal. For example, in a coder design application, we must first completely characterize the auditory Ψ-domain signal-to-noise ratio (e.g., the ratio of loudness to loudness noise), and then describe it in terms of the Φ-domain signal-to-noise ratio (i.e., the ear canal pressure or intensity and the intensity JND_I).

It is highly recommended that the serious reader carefully study chapters 10, 11, and 13, and Appendix D of Yost (1994). Other important general sources are (Gelfand 1981) and the excellent book of (Littler 1965), which along with (Fletcher 1995), was a primary source of

information for the author.

Some history. By 1918 AT&T had decided that if they were going to transmit speech across the country by wire, it was essential that they understand the physical processes behind auditory communication. This was clearly articulated in an internal report written in 1920 by J.Q. Stewart of the AT&T Development and Research (D&R) Department which oversaw the funding for the Western Electric Engineering group:

> The desirability of making direct studies of the physics and psychology of hearing in the quality investigation is becoming increasingly more evident. Research on the physical nature of speech alone will not be sufficient to establish the physical basis for the prediction of articulation; but it must be supplemented by studies of hearing. Indeed, the latter seem to be even more important, because the progress which already has been made toward the formulation of a general philosophy of articulation has not been dependent to any degree on knowledge of the physical characteristics of speech, but has been dependent on hypothesis relating to hearing.

This set the tone for the next 27 years of work. By 1922 a research project was in full gear due to the work of Harvey Fletcher, who had joined AT&T's Western Electric engineering department in 1916. Fletcher's Ph.D. concerned what is now known as the "Millikan Oil drop experiment," and AT&T hired him, based on his basic understanding of the physics of the electron, to help build a better telephone system. Fletcher quickly became the leader of this large and important research effort, and by 1924 AT&T was fully committed to funding basic research on speech perception. This was a true team effort, led by Fletcher.

Fletcher and Wegel (1922) accurately measured (for the first time) the threshold of hearing. The year after their study, Knudsen, a student of Fletcher's, was the first to measure the pure-tone intensity JND (Knudsen 1923). Then Fletcher (Fletcher 1923a), with the help of Wegel and Lane (Wegel and Lane 1924) provided critical and detailed tone-on-tone masking data. Fletcher and Steinberg then showed the relation between Wegel and Lane's masking data and partial loudness (Fletcher and Steinberg 1924). In 1924 Wegel and Lane outlined the physical theory of the cochlear traveling wave. Kingsbury measured iso-loudness contours which defined the *loudness-level*, or *phon* scale (Kingsbury 1927; Fletcher 1929, (p. 227)). Finally, Riesz (Riesz 1928) conducted an extensive study of tonal masking for probes near the masker frequency (Littler 1965; Yost 1994).

This series of theoretical and experimental studies had shown that cochlear filtering results from a traveling wave on the basilar membrane, and the power-law nature of loudness growth in the cochlea was established. Fletcher summarized this work in the first book to be written on speech and hearing (Fletcher 1929), bringing him world-wide acclaim.

In 1933, based on detailed loudness and masking data of Munson, and a theory worked out by Fletcher, Fletcher and Munson published the first model of loudness (Fletcher and Munson 1933). They described how the auditory signal is broken down into many frequency bands and compressed by the cochlea prior to being neurally coded. They described partial loudness as the neural rate, provided the functional relationship between masking and partial loudness,

and show how partial loudness is summed to given the total loudness (Allen 1996).[1]

By 1935 it was proposed that the randomness in the neural representation is what makes the Ψ variables random variables (Montgomery 1935; Stevens and Davis 1938). Based on the nature of hearing-loss and loudness recruitment, it became clear by 1937 that the basilar membrane signal is compressed, that this compression is lost when the cochlea is damaged (Steinberg and Gardner 1937), that the cochlear damage is due to the loss of hair cells (Lorente de No 1937; Carver 1978), and that the resulting loss of outer hair cells causes loudness recruitment (Allen 1996). At high intensities, all the neurons are stimulated and the loudness is approximately the same as in the normal ear; however, because of the loss of the nonlinear compressive action in the recruiting ear of the outer hair cells, the dynamic range of the cochlea is reduced in the recruiting ear.

By 1947 the similarity between the JND and masking had been clearly articulated and quantified (Miller 1947). However, the significance of the neural noise proposal of Montgomery was not appreciated for at least 30 years, when Siebert applied signal detection theory to various JND data (Siebert 1965). As will be described below, while neural noise appears to determine the limits of the pure-tone JND task under many conditions, it may not the limiting factor in all listening tasks. If another uncertainty (e.g., external noise) dominates the neural uncertainty, then that factor will limit the JND.

1.2 Summary

Fletcher's 1933 loudness model is basic to our present understanding of loudness coding. This model had clearly described tonal loudness in terms of an additive neural rate code. While we still need to fill in many details, in my opinion this basic idea (however controversial), is correct. By 1950 it had been extensively tested (Fletcher and Munson 1937; Fletcher 1938a; Munson and Gardner 1950; Fletcher 1995).

From the loudness model, several predictions are important. These include the auditory threshold, which is modeled as zero loudness (Fletcher and Munson 1933), the JND, which provides an inferred measure of the loudness uncertainty (i.e., "internal noise") (Montgomery 1935; Miller 1947; Hellman and Hellman 1990; Allen and Neely 1997), the masked threshold, which is a generalized JND measure of signal uncertainty (Fletcher and Munson 1933; Allen and Neely 1997), signal duration effects (Munson 1947), and the critical band (Fletcher 1938a; Fletcher 1938b). Many other phenomena (e.g., the frequency JND, beats) appear to be accounted for by these models (Fletcher 1995), but the details need to be reevaluated in terms of the theory of signal detection (TSD).

What is seriously lacking in our present day understanding of auditory signal processing is a quantitative model of masking. While Fletcher's loudness model shows us how to calculate the loudness given the masked threshold (Fletcher and Munson 1933; Fletcher and Munson 1937; Fletcher 1938a; Fletcher 1940), and provides detailed experimental results on the relations between masking and loudness of tones and noise of various bandwidths (Fletcher 1995), he did not (nor did Zwicker) provide us with an accurate method for the direct calculation of the masking pattern for arbitrary signals. Reasons for this include the lack of a quantitative

[1] Unfortunately, while this paper was highly regarded, the information was not widely digested and accepted by the research community. For example, the idea that tonal loudness is additive is still controversial (Marks 1979).

understanding (i.e., models) of (a) the highly nonlinear, upward spread of masking (Wegel and Lane 1924), (b) the stochastic nature of masking (Montgomery 1935), and (c) beats.[2] Thus masking models are the key to understanding hearing, improving speech coding, and quantifying models of loudness. Masking is one of the most important topics in psychophysics, yet it is arguably the most poorly understood.

During the 1960's the theory of signal detection (TSD) was identified as an important tool for quantifying the detection of probe signals. What is missing is detailed models of the decision variables used by the central nervous system (CNS). For the JND case, where the probe is a copy of the masker, the relevant decision variable is the change in loudness (Allen 1996; Neely and Allen 1996; Hellman and Hellman 1990). In the case of masking, where the probe and masker are different, the decision variable is unknown and must be experimentally deduced with a model and experimental data for a multiplicity of maskers and probes. To model the masked threshold (i.e., masking) we need more than a model of loudness, we need a model of the probe signal-detection task. This model must start with the partial loudness of the masker plus probe, and describe the detection probability of the probe signal, as a function of the probe and masker physical parameters.

1.3 Overview of this article.

In the next section we will define important concepts such as loudness, the intensity JND (sometimes called "self-masking"), masking, and critical bands. The discussion in that section is limited to describing the definitions. Readers familiar with these definitions may move directly to the analysis and discussion in the subsequent sections on **The loudness SNR** Section 3, **Narrowband maskers** Section 4, and use Section 2 as a reference.

We shall argue that masking is synonymous with the uncertainty of the Ψ representation. In engineering jargon, masking results from the "quantizing noise" of the loudness representation. Because of this relationship, and its simplicity, the JND has special significance to the theory of masking. To analyze the problem in more detail we need to make some distinctions about n-interval force choice (n-IFC) methods, intensity modulation detection, pure-tone and "frozen-noise" maskers of various bandwidths, and forward masking. Finally we describe a time-domain nonlinear model of the cochlea and auditory system that explains these data and discuss how one might model masking by arbitrary signals, such as music and speech.

When representing signals in a computer there are two basic models: fixed-point and floating-point. In the telephone industry there is a third standard called μ-law. In a fix-point representation the noise is fixed to 1/2 of the least significant bit (LSB). Thus the noise is, to a first-order approximation, independent of the signal level, and the signal-to-noise ratio is proportional to the signal. In a floating-point representation, the noise is a fixed percentage of the signal level. For example, with an 8 bit mantissa, the noise floor would be approximately 8×6=48 dB below the signal. Thus the SNR is roughly independent of the signal level. In a μ-law signal the noise depends on the signal level. The ideal μ-law device is similar to a logarithmic compression function and provides a floating-point signal representation, with a 38

[2] Fletcher was also busy working out his empirical theory of speech perception (Fletcher 1922; Fletcher and Galt 1950; Allen 1994).

dB dynamic range, over the signal range from -40 to 0 dB, and a fixed-point representation with a constant noise level over the signal range from -80 to -40 dB (Jayant and Noll 1984).

We shall see that in the auditory system, the representation noise is a function of the signal level in a way that is similar to μ-law coding. At low intensities the loudness SNR improves, and at higher intensity it saturates, with a maximum of 30 to 40. Because of the compression of the loudness, specified by the loudness power law, the ratio of the loudness SNR to the signal SNR equals the exponent in the loudness power law. For wideband signals at moderate to high intensities the compression is approximately the square root of the pressure (fourth root of intensity). This means that a loudness SNR of ≈40, referred to the input signal domain, is ≈10 [i.e., 10 $\log_{10}(10) = 10$ dB].

2 Definitions

The definitions in this section are summarized in Tab. reftab:defs.

2.1 Loudness

Loudness ($\mathcal{L}(I)$) is the name of the Ψ intensity corresponding to an acoustic Φ signal. One of the major conclusions of modern psychophysics is that Ψ variables are random variables (or processes). We define the *single-trial loudness* $\widetilde{\mathcal{L}}$ as the Ψ random processes that represents the loudness corresponding to each presentation of the signal (the $\tilde{}$ is used to indicate a random variable; all Ψ variables are represented in calligraphic font). A *trial* is defined as a stimulus *presentation*, followed by a subject *response*. The presentation can be a single *interval*, as in loudness scaling methods, or multiple presentations (e.g., an ABX is an example of a 3IFC method. Signals A and B are first presented, followed by X, which is either A or B.) (Yost 1994). Since the Ψ variables are random, the subject responses must be random, with a degree of randomness (variance) that depends on the task.

The expected value of $\widetilde{\mathcal{L}}(I)$,

$$\mathcal{L}(I) \equiv \mathcal{E}[\widetilde{\mathcal{L}}(I)]$$

, characterizes the loudness $\mathcal{L}(I)$ at intensity I, while the variance

$$\sigma^2_{\mathcal{L}}(I) \equiv \mathcal{E}[[\widetilde{\mathcal{L}}(I) - \mathcal{L}(I)]^2],$$

characterizes the loudness JND $\Delta\mathcal{L} = d'\sigma_{\mathcal{L}}$, where the proportionality factor d' depends on the subject's criterion and on the experimental design. Loudness is an important example of an auditory Ψ variable; without a model of loudness, it is *not* possible to relate (i.e., model) many different experimental results, because the subject's responses are a function of the loudness. A model of the loudness of simple sounds, such as tones and narrow bands of noise, is critical to the theory of masking because the masking is directly related to the *standard deviation* (s.d.) $\sigma_{\mathcal{L}}$.

An example. Without loudness we cannot make Ψ models of the signals we are processing. An example should help clarify this point. In 1947 Munson studied the question of the effects of signal duration on tonal loudness. He matched the loudness of a reference tone at frequency

$f_{\text{ref}} = 1$ kHz, duration $T_{\text{ref}} = 1$ second, and intensity $I^*_{ijk}(I_i, f_j, T_k)$, to target tones of intensities I_i = (30 – 90) dB SPL, frequencies f_j = (0.25 – 10) kHz, and durations T_k = (5 – 500) ms. For example, if $\mathcal{L}(I, f, T)$ is the loudness of a tone at intensity I, frequency f, and duration T, then the *loudness-level* I^*_{ijk}, in phons, is defined by the relation

$$\mathcal{L}(I^*_{ijk}, f_{\text{ref}}, T_{\text{ref}}) = \mathcal{L}(I_i, f_j, T_k).$$

Munson was able to model these experimental results accurately by first transforming the matched intensity I^*_{ijk} to a *loudness-rate* $\mathcal{L}_t(t)$ per unit time, using the power-law relation of Fletcher and Munson (Fletcher and Munson 1933), and then integrating the results with a lowpass filter $E_s(t)$, which he called the *sensation integral*. The form of $E_s(t)$ was determined from the data so as to best fit all his experimental results. He then *predicted* the loudness matches to tones having duration of up to one minute measured previously by von Bekesy. The key to this model is the nonlinear transformation to the loudness variable, and an integration of the resulting loudness-rate over time. The additivity of loudness between ears and across frequency had previously been demonstrated in (Fletcher and Munson 1933), and is a key property of these loudness models.

2.1.1 Critical band

The threshold of a pure tone in wideband noise is determined by the loudness standard deviation $\sigma_{\mathcal{L}}$ and the bandwidth of the filters, as measured by the *equivalent rectangular bandwidth* (ERB), or critical bandwidth. The *critical band* is the name Fletcher gave to the bandpass filters of the cochlea. When a tone centered on the center frequency of a bandpass filter is presented in a wide band of noise, the filter removes the noise outside of the passband of the filter. Thus the loudness SNR at the output of the filter depends critically on the bandwidth of the filter. This means that the detectibility of the tone in noise is determined by the filter's equivalent power bandwidth (ERB). Since masking is the increase in the threshold of a tone, due to the presence of noise, the critical bandwidth plays an important role in masking when using wideband maskers.

2.1.2 Modeling loudness

The basic transformation to loudness is a several-step process. First, the input signal is filtered by the cochlea. Second, the dynamic range of the signals is compressed by the outer hair cells on the basilar membrane (Allen and Neely 1992; Allen 1996; Neely and Allen 1996). Third, the signals are encoded by the inner hair cells and neurons, and a stochastic neural representation results. Masking and partial loudness are formed from stochastic neural representation. There is extensive evidence that these two quantities are functionally related (Fletcher 1995; Allen 1995). Finally the neural representation is processed by the nervous system and the total loudness is evaluated. It was proposed by Fletcher and Munson in 1933, for simple signals such as tones and noise, that this final processing is a simple sum of the neural rate. While this assumption of partial loudness additivity has been controversial, the assumption of additivity has held up amazingly well (Marks 1979). An analogy that seems appropriate is Newton's apple. Newton could never prove that the apple would always fall, but it always did. Only

with the discovery of quantum mechanics, were the important limits of Newton's law $F = ma$ uncovered. It is probably true that partial loud does not always add. But under the conditions of these simple experiments with tones, it always does. Thus it is important to appreciate both the limitations and the power of the additivity of partial loudness (Allen 1995).

The two models that have described the above steps are those of Fletcher and Zwicker. Both of these models are described in the frequency domain. Both models are deterministic. Loudness however is a random variable. We shall show that after a transformation of tone and noise intensity JND data into the loudness domain, the JND data are greatly simplified.

The emphasis in this article is on predicting masking for arbitrary signals. To do this we need a time-domain loudness model. The problem with all loudness models (including the present one) is the lack of detailed understanding, and therefore specification, of the transformation between the ear canal pressure and the motion of a point on the basilar membrane. This transformation has two components: a linear filtering component and a compressive (nonlinear) component. While approximate solutions to this problem have been proposed, there is presently no accepted model of the compressive transformation. Until such a theory is more fully developed, we must continue to deal directly with experimental data (Allen 1991). The greatest simplification and understanding of these experimental data is found in a description of the data in terms of *masking patterns* and *partial loudness excitation patterns*. As we have refined our understanding of the nonlinear excitation pattern model, we have been able to account for diverse types of experimental data. However, we cannot consider this problem solved until the physics of the nonlinear transformation, involving the outer hair cells of the cochlea, is fully described.

2.1.3 Loudness growth

Φ *intensity* is power per unit area. *Loudness*, in *sones* or *loudness units* (LU),[3] is the name commonly given to the Ψ intensity. When there are standing waves in the ear canal, the ear canal pressure is a sum of both inward- and outward-traveling pressure waves. It seems reasonable, but has not been adequately proven, that the power flow *into* the ear should be a better measure of hearing performance than the total pressure. Loudness depends in a complex manner on a number of acoustical variables, such as intensity, frequency, and spectral bandwidth, and on the temporal properties of the stimulus, as well as on the mode of listening (in quiet or in noise, binaural or monaural stimulation). Isoloudness contours describe the relation of equal loudness between tones or between narrow bands of noise at different frequencies.

In 1924 Fletcher and Steinberg published an important article on the measurement of the loudness of speech signals (Fletcher and Steinberg 1924). In this paper, when describing the growth of loudness, the authors state

> the use of the above formula involved a *summation of the cube root of the energy rather than the energy*.

This cube-root dependence was first described by Fletcher the year before (Fletcher 1923a). Today any power-law relation between the intensity of the physical stimulus and the psy-

[3] Sones and LU are related by a scale factor: 1 Sone is 975 LU.

chophysical response is referred to as *Stevens's law* (Rosenblith 1959; Yost 1994). Fletcher's 1923 loudness growth equation established the important special case of loudness for Stevens's approximate, but more general, psychological "law."

Cochlear nonlinearity: How? What is the source of Fletcher's cube root loudness growth (i.e., Stevens's law)? Today we know that the basilar membrane motion is nonlinear, and that cochlear outer hair cells (OHC) are the source of the basilar membrane nonlinearity and, as a result, the cube root loudness growth observed by Fletcher.

From noise trauma experiments on animals and humans, it is now widely accepted that recruitment (abnormal loudness growth) occurs in the cochlea (Carver 1978). In 1937 Lorente de No theorized that recruitment is due to hair cell damage (Lorente de No 1937). Animal experiments have confirmed this prediction and have emphasized the importance of outer hair cells (OHC) loss (Liberman and Kiang 1978; Liberman and Dodds 1984). This loss of OHC's causes a loss of the basilar membrane compression first described by Rhode in 1971 (Allen and Fahey 1992; Yost 1994; Pickles 1982, (p. 291)). It follows that the cube-root loudness growth results from the nonlinear compression of basilar membrane motion due to stimulus-dependent voltage changes within the OHC.

We still do not know precisely what controls the basilar membrane nonlinearity, although we know that it is related to outer hair cell length changes which are controlled by the OHC membrane voltage (Santos-Sacchi and Dilger 1987). This voltage is determined by shearing displacement of the hair cell cilia by the tectorial membrane. We know that the inner hair cell (IHC) has a limited dynamic range of less than 60 dB, yet it is experimentally observed that these cells code a dynamic range of about 120 dB (Allen 1996). Nonlinear compression by cochlear OHCs, prior to IHC detection, increases the dynamic range of the IHC detectors. When the OHCs are damaged, the compression becomes linear, and loudness recruitment results (Steinberg and Gardner 1937).

2.1.4 Loudness additivity

Fletcher and Munson (1933) showed, for tonal stimuli, *(1)* the relation of iso-loudness across frequency (loudness-level in phons), *(2)* the dependence of loudness on intensity *(3)* a model showing the relation of masking to loudness, and *(4)* the basic idea behind the critical band (critical ratio).

Rather than thinking directly in terms of loudness growth, they tried to find a formula describing how the loudnesses of several stimuli combine. From loudness experiments with low- and highpass speech and complex tones (Fletcher and Steinberg 1924; Fletcher 1929) and from other unpublished experiments over the previous 10 years, they found that loudness adds. Today this model concept is called *loudness additivity*. Their hypothesis was that when two equally loud tones that do not mask each other are presented together, the result is "twice as loud." They showed that N tones that are all equally loud, when played together, are N times louder, for N up to 11, as long as they do not mask each other. Fletcher and Munson found that loudness additivity held for signals between the two ears as well as for signals in the same ear. When the tones masked each other (namely, when their masking patterns overlapped), additivity still held, but over an attenuated set of patterns (Fletcher and Munson 1933), since

the overlap region must not be counted twice. This 1933 model is fundamental to our present understanding of auditory sound processing.

The argument. Let $G(p_1, p_2)$ be the nonlinear compression function that maps the ear canal pressure p_1 at frequency f_1 and p_2 at f_2 into the loudness in sones, *under the condition that the tones are far enough apart in frequency that they do not mask each other*. When one tone masks another, the loudness $\mathcal{L}$ is always less than G (i.e., masking always reduces the loudness). When each tone is presented alone, there is no masking, so $\mathcal{L} = G$. It also follows that $\mathcal{L}_1 = G(p_1, 0)$ and $\mathcal{L}_2 = G(0, p_2)$. We assume that $G(0, 0) = 0$ and $G(p_{\text{ref}}, 0) = 1$, where p_{ref} is either 20 μPa or the threshold of hearing at 1 kHz. The problem is to find $G(p_1, p_2)$.

Step 1. The pressure p_1 is taken as the reference level for the experiment with $f_1 = 1$ kHz. The level of pressure p_2, at frequency f_2, is next determined by requiring that its loudness be equal to that of p_1. We call this pressure $p_2^*(p_1, f_2)$, since it is a function of both p_1 and f_2. In terms of the compression function G, p_2^* is defined by

$$G(0, p_2^*) = G(p_1, 0). \tag{1}$$

Step 2. Fletcher and Munson scaled the reference pressure p_1 by scale factor α and defined α^* such that the loudness of $\alpha^* p_1$ is equal to the loudness of p_1 and p_2^* played together. In terms of G this condition is

$$G(\alpha^* p_1, 0) = G(p_1, p_2^*). \tag{2}$$

This equation defines α^*.

Results. For f_1 between 0.8 and 8.0 kHz, and f_2 far enough away from f_1 (above or below) so that there is no masking, $20 \log_{10}(\alpha^*(I))$ was found to be ≈9 dB for p_1 above 40 dB-SL. Below 40 dB-SL, this value decreased linearly to about 2 dB for p_1 at 0 phons, as shown in Fig. 1. It was found that the loudness $G(p_1, p_2^*)$ does not depend on $p_2^*(p_1, f_2)$ as f_2 is varied. Thus we may write $\alpha^*(p_1, p_2^*)$ to show its dependence on p_1 and its independence of p_2^*.

Fletcher and Munson found an elegant summary of their data. They tested the assumption that

$$G(p_1, p_2) = G(p_1, 0) + G(0, p_2), \tag{3}$$

namely that the loudnesses of the two tones add. Using Eq. 1, Eq. 3 becomes

$$G(p_1, p_2^*) = 2G(p_1, 0). \tag{4}$$

Combining Eq. 2 and Eq. 4 gives the nonlinear difference equation

$$G(\alpha^*(p_1) p_1, 0) = 2G(p_1, 0), \tag{5}$$

which determines G once $\alpha^*(p_1)$ is specified. $G(p)$ may be found by graphical methods, or by numerical recursion, as shown in Fig. 136 (Fletcher 1995, Page 190).

From this formulation Fletcher and Munson found that at 1 kHz, and above 40 dB SPL, the pure-tone loudness G is proportional to the cube root of the signal intensity [$G(p) = (p/p_{\text{ref}})^{2/3}$,

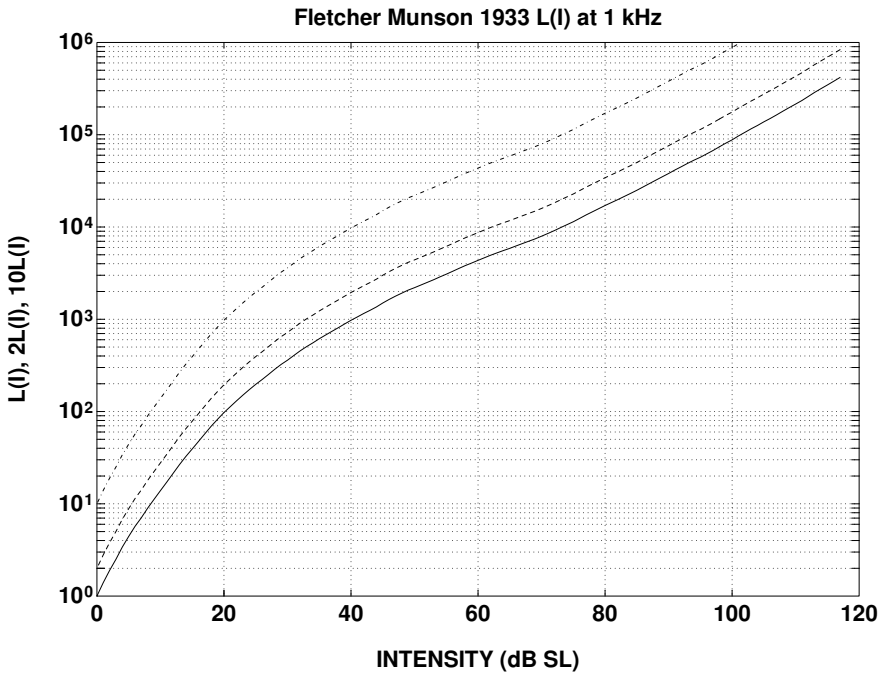

Figure 1: This figure shows the loudness growth $\mathcal{L}(I)$ (solid line) from (Fletcher 1995) in LU (975 LU is 1 Sone), along with $2\mathcal{L}(I)$ (dashed line) and $10\mathcal{L}(I)$ (dot-dashed line) for reference. To determine $\alpha^*(I)$ draw a horizontal line that crosses the $2\mathcal{L}(I)$ and $\mathcal{L}(I)$ curves, and note the two intensities. The dB-difference is $20\log_{10}(\alpha^*(I))$. For example, the total loudness of two 40 dB SL tones presented to each of the two ears is 2000 LU (2 sones), and is equal to the loudness of a single tone at 49 dB SPL. Thus α^* at 40 dB is 9 dB.

since $\alpha^* = 2^{3/2}$, or 9 dB]. This means that if the pressure is increased by 9 dB, the loudness is doubled. Below 40 dB SPL, loudness was frequently approximated as being proportional to intensity [$G(p) = (p/p_{\text{ref}})^2$, $\alpha^* = 2^{1/2}$, or 3 dB]. Figure 1 shows the loudness growth curve. Estimated values of $\alpha^*(I)$ are given in (Fletcher 1995, Table 31, page 192).

2.2 The just-noticeable difference in intensity

Basic to psychophysics, more fundamental than masking, is the concept of a *just-noticeable difference* (JND), or *difference limen* (DL). A primary premise of the *auditory theory of signal detection* (ATSD) is that the JND is a measure of the Ψ uncertainty (noise) (Yost 1994). That is, if we increase the intensity from I to $I + \Delta I$, such that we can just hear the change in intensity, then ΔI should be proportional to the uncertainty of the Ψ representation of I. ΔI can also have a component due to the signal uncertainty, called "external" noise, as when the signal is "roved," for example (Green 1988b). This idea is captured by the equation (Green 1988b, p. 113) $\Delta I = d'\sigma_I$, where ΔI is the intensity JND, and σ_I^2 is the variance of I due to the internal noise,*reflected back to the input of the system.* This equation says that the just-noticeable perceptual change in intensity is proportional to the internal noise, as reflected in the intensity variance. This equation defines a model for the detection of the change in signal intensity. While this model is widely accepted, it has not been quantitatively verified (Allen and Neely

Table 1: Table of definitions.

Abbreviation	Symbol	Definition
Intensity	I	Eq. 6 or $\mathcal{E}[\widetilde{I}]$
Intensity detector output	$\widetilde{\mathcal{I}}$	Eq. 10
Single-trial loudness	$\widetilde{\mathcal{L}}$	
Loudness	$\mathcal{L}$	$\mathcal{E}[\widetilde{\mathcal{L}}]$
Loudness s.d.	$\sigma_{\mathcal{L}}$	$\mathcal{E}\{[\widetilde{\mathcal{L}}(I) - \mathcal{L}(I)]^2\}$
Intensity s.d.	σ_I	$\sigma_{\mathcal{L}} / \left\lvert \frac{d\mathcal{L}}{dI} \right\rvert$
JND$_I$	ΔI	Intensity JND
JND$_{\mathcal{L}}$	$\Delta\mathcal{L}$	Loudness JND
SNR$_I$	I/σ_I	Intensity SNR
SNR$_{\mathcal{I}}$	$\mathcal{I}/\sigma_{\mathcal{I}}$	Intensity detector SNR
SNR$_{\mathcal{L}}$	$\mathcal{L}/\sigma_{\mathcal{L}}$	Loudness SNR
d-prime	d'	Eq. 11
Weber Fraction	J	$\Delta I/I$
Probe gain	α	Probe scale factor
Decibel level	β	$10\log_{10}(I/I_{\text{ref}})$
Log-loudness	$\mathcal{L}_{\log}$	$10\log_{10}(\mathcal{L})$
Loudness exponent	γ	$d\mathcal{L}_{\log}/d\beta$
	$\widetilde{\Delta\mathcal{I}}(t,\alpha)$	$\widetilde{\mathcal{I}}(t,\alpha) - \widetilde{\mathcal{I}}(t,0)$
	$\Delta\mathcal{I}(t,\alpha)$	$\mathcal{I}(t,\alpha) - \mathcal{I}(t,0)$
Threshold probe gain	α^*	$d'(\alpha^*) = 1$

1997).

Definition of $I(t)$. The *intensity* of a sound (in watts/cm^2) is a physical variable, defined as the square of the pressure divided by the acoustic impedance $I = P^2/\rho c$ where P is the root-mean-square (RMS) pressure, ρ is the density of air, and c is the sound speed. In the time-domain when the impedance is fixed, it is common to define the *intensity* in terms of the time-integrated squared signal pressure $p(t)$, namely

$$I(t) = \frac{1}{\rho c T} \int_{t-T}^{t} p^2(t) dt \tag{6}$$

These two alternative definitions of intensity differ by the integration over and normalization by a fixed interval T seconds long. When the signal $p(t)$ is deterministic, as in the case of pure tones, we shall define T to be the duration of the tone. When the signal is a Gaussian noise, we let $p(t) = \tilde{n}(t)$ [i.e., $\mathcal{N}(0, \sigma_n)$], and T be the duration of the noise, leading to $I \equiv \mathcal{E}[\widetilde{I}(t)] = \sigma_n^2$.

Definition of ΔI. An *increment in sound intensity* $\Delta I(\alpha)$ may be defined in terms of a positive pressure scale-factor $\alpha \geq 0$ applied to the signal pressure $s(t)$, as

$$I(t,\alpha) = I(t,0) + \Delta I(t,\alpha) \quad (7)$$

$$= \frac{1}{\rho c T}\int_{t-T}^{t}[s(t)+\alpha s(t)]^2 dt. \quad (8)$$

Expanding this relation

$$\Delta I(t,\alpha) = (2\alpha+\alpha^2)I(t,0). \quad (9)$$

The estimate of the intensity $I(t,\alpha)$ is a function of the time and the signal gain, and $I(t,0)$ indicates the case where $\alpha = 0$.

Definition of an intensity detector. In psychophysics the ear is frequently modeled as an intensity detector. It is useful therefore to introduce this popular model formally, and compare its performance with that of the ear. We define an *intensity detector* as the intensity $I(t)$ plus the *internal noise* of the detector,

$$\widetilde{\mathcal{I}}(t) \equiv I(t) + \tilde{\epsilon}(t) \quad (10)$$

which is modeled as an independent Gaussian random process $\tilde{\epsilon}(t)$, having zero-mean and variance σ_ϵ^2 [i.e., $\mathcal{N}(0,\sigma_\epsilon)$]. The internal noise limits the JND for nonrandom signals, such as pure tones.

Throughout this article, we shall only consider zero-mean signals [e.g., $s(t)$] when using the above definition of the intensity detector. One must carefully distinguish I, the observable intensity, and $\widetilde{\mathcal{I}}$, a *decision variable* that is not observable. If we think of the energy detector as a crude model of the ear, $\widetilde{\mathcal{I}}$ is the decision variable which represents the Ψ intensity (i.e., the loudness). We will show that the intensity detector is *not* a good model of the ear, because both its level dependence and its internal noise are vastly different from those of the ear [i.e., $\mathcal{L}(I) \neq \mathcal{I}(I)$ and $\sigma_{\mathcal{L}}(I) \neq \sigma_{\mathcal{I}}(I)$] (Allen and Neely 1997). However, the intensity detector is an important "straw man" candidate for comparison purposes. It is important to remember that the auditory brain has no access to the intensity of the stimulus. It only receives auditory information from the auditory nerve.

The *mean detector intensity* is defined as $\mathcal{I} \equiv \mathcal{E}[\widetilde{\mathcal{I}}]$ and the *variance of the detector intensity* is defined by $\sigma_{\mathcal{I}}^2(I,T) \equiv \mathcal{E}[\widetilde{\mathcal{I}}^2] - \mathcal{I}^2$. From this definition, $\mathcal{I} = I$ and $\Delta\mathcal{I}(\alpha) = \Delta I(\alpha)$. The variance represents the uncertainty of the internal decision variable and plays a fundamental role in the theory of signal detection. We shall see that $\sigma_{\mathcal{I}}$ is a function of both the mean intensity and the duration, but for deterministic signals is simply equal to the internal noise of the energy detector [i.e., $\widetilde{\mathcal{I}}(t,\alpha]$ is $\mathcal{N}(I(\alpha),\sigma_\epsilon)$). For stochastic signals $\sigma_{\mathcal{I}}(I, T\to\infty) = \sigma_\epsilon$, namely as the duration of the random signals is increased, $\sigma_{\mathcal{I}}(I,T)$ is finally limited by the internal noise. This means there are conditions (e.g., large T or small I) where the internal noise of the detector will dominate its performance.

2.2.1 Definition of the JND$_I$

The *just-noticeable difference in intensity* (JND$_I$) is determined by finding that value of α, which we call α^* (the $*$ is used to indicate the particular value of α), such that the subject

can correctly identify the decision variable $\widetilde{\mathcal{I}}(t,\alpha^*)$ from $\widetilde{\mathcal{I}}(t,0)$ 50% of the time, adjusted for chance. It is convenient and common to refer to JND$_I$ as simply $\Delta I(I)$ rather than using the more cumbersome (but more precise) composite-function notation $\Delta I(\alpha^*(I))$. The symbol $\Delta I(I)$ can be confusing because one does not know if it means JND$_I \equiv \Delta I(\alpha^*(I))$, $\Delta I(\alpha)$ (i.e., the change in intensity corresponding to some gain α), or $\Delta\mathcal{I}$. For this reason JND$_I$ and JND$_\mathcal{I}$ are more precise symbol than the widely used $\Delta I(I)$.

For the intensity detector defined above, one may analytically determine JND$_I$ and show that $\Delta I(\alpha^*) = \sigma_\epsilon$. For human subjects we must run an experiment, such as a 2-IFC comparison, and make a model of the observations. In this case the value of α that satisfies the 50% above-chance discrimination condition, $\alpha^*(I)$, depends on I (i.e., $\Delta I/I$ depends on I for tones, but is approximately constant for wideband noise).

2.2.2 Weber's law.

The intensity JND is frequently expressed as a *relative JND*$_I$ called the *Weber fraction*, defined by $J(\alpha^*) \equiv \Delta I/I$. Weber's law $J(I)$, that J is independent of I, was first proposed in 1846 (Weber 1988).

Weber's law and pure tones. The Weber fraction $J(\alpha^*(I))$ *is* a function of intensity for the most elementary signal, the pure tone (Knudsen 1923; Riesz 1928; Jesteadt et al. 1977). This observation is referred to as the *near miss to Weber's law* (McGill and Goldberg 1968). The "near miss" shows that the ear is not an energy detector, since for an energy detector $\Delta I = d'\sigma_\epsilon$ is independent of intensity. For recent discussions of why Weber's law holds approximately for tones (Green 1988a; Green 1970, page 721), or why it holds for wide-band noise more than 20 dB above threshold, we refer the reader to the detailed discussion of these questions by Viemeister (1988), Hartmann (1997), and (Allen and Neely 1997).

The JND for an energy detector. Next we review the derivation of the JND for the energy detector. Two independent signals [either $s(t)$ or $\tilde{n}(t)$, $\mathcal{N}(0,\sigma_n)$] are presented to the energy detector with gains $\alpha = 0$ and α, having the decision variable $\widetilde{\mathcal{I}}(t,\alpha)$. We would like find α such that the more intense signal is greater than the less intense signal 75% of the time. This task is equivalent to the following: Find α such that the difference $\widetilde{\Delta\mathcal{I}}(t,\alpha) \equiv \widetilde{\mathcal{I}}(t,\alpha) - \widetilde{\mathcal{I}}(t,0)$ is greater than zero 75% of the time. It is assumed that $\widetilde{\Delta\mathcal{I}}$ is Gaussian with a variance of $2\sigma_\mathcal{I}^2$ [the variances $\sigma_\mathcal{I}^2(0)$ and $\sigma_\mathcal{I}^2(\alpha)$ are assumed to be similar enough to be treated as equal].

When $\Delta\mathcal{I}(\alpha) \equiv \mathcal{E}[\widetilde{\Delta\mathcal{I}}(t,\alpha)]$ is $\sigma_\mathcal{I}$ [i.e., when $\widetilde{\Delta\mathcal{I}}(t,\alpha)$ is $\mathcal{N}(\sigma_\mathcal{I},\sqrt{2}\sigma_\mathcal{I})$], the probability that $\Delta\mathcal{I} > 0$ is ≈ 0.76. This probability is close enough to the definition of 0.75 that it has been adopted as the *de facto* standard detection threshold (Green 1988b). The ratio of $\Delta\mathcal{I}$ to $\sigma_\mathcal{I}$ is an important statistic of the experimental signal uncertainty called d'"

$$d'(\alpha) \equiv \frac{\Delta\mathcal{I}(\alpha)}{\sigma_\mathcal{I}(\alpha)}. \tag{11}$$

Using this notation, the definition of α^* is $d'(\alpha^*) \equiv 1$.

Thus with the assumption of an intensity detector having Gaussian detection variables of equal variance, and a detection criterion of 76%, the Weber fraction is[4]

$$J = \frac{\sigma_{\mathcal{I}}(I)}{\mathcal{I}}. \tag{12}$$

The ratio of the intensity to the intensity variance defines an intensity SNR $\text{SNR}_{\mathcal{I}} \equiv \mathcal{I}/\sigma_{\mathcal{I}}$, which is more intuitive than the Weber fraction.

The gain α^* is then determined from Eq. 9 and Eq. 12

$$2\alpha^* + (\alpha^*)^2 = \frac{1}{\text{SNR}_{\mathcal{I}}}, \tag{13}$$

and since $\alpha^* \geq 0$, we have $\alpha^* = \sqrt{1 + 1/\text{SNR}_{\mathcal{I}}} - 1$. This last equation allows one to calculate α^* given $\text{SNR}_{\mathcal{I}}$, or estimate $\text{SNR}_{\mathcal{I}}$ given a measurement of α^* from a pure tone intensity JND experiment.

Internal versus external noise. There are two commonly identified types of masking noise, *internal* and *external* noise (Buus 1990). Both of these two types of noise are modeled as Ψ (e.g., loudness) variability, which is synonymous with masking. Internal noise is due to the stochastic nature of the Ψ representation, while external noise is due the stochastic nature of the Φ representation (i.e., variability in the stimulus), which is transformed into Ψ variability by the auditory system. Internal noise sets the fundamental limit on the representation of the auditory system. Roving the signal is a technique designed to make the external noise dominate.

For the case of external noise, it is possible to show (M.M. Sondhi, personal communication) that

$$\sigma^2_{\mathcal{I}}(I,T) = \frac{2I^2}{T^2B^2} + \sigma^2_{\epsilon}, \tag{14}$$

where B is an effective bandwidth (that depends on T) defined by

$$B(T) \equiv \left(\int_{\tau=0}^{T} \int_{t=0}^{T} \rho^2(t-\tau) dt d\tau \right)^{-\frac{1}{2}}$$

and $\rho(t-\tau) \equiv \mathcal{E}[\tilde{n}(t)\tilde{n}(\tau)]/I$ is the normalized [i.e., $\rho(0) = 1$] covariance of the stochastic signal $s(t) = \tilde{n}(t)$. Thus for the intensity detector with a Gaussian input having a variance that dominates the detector noise, Weber's law holds, and $J(I) = \sqrt{2}/TB$, or $\text{SNR}_{\mathcal{I}} = TB/\sqrt{2}$. The product TB is called the *degree-of-freedom* parameter.

2.2.3 Definition of $\Delta\mathcal{L}$.

Any superthreshold increment in the sound intensity must have a corresponding loudness increment. A *loudness increment* $\Delta\mathcal{L}(I)$ is defined as the change in loudness $\mathcal{L}(I)$ corresponding to an intensity increment $\Delta I(I)$. When $\Delta I(I)$ is the JND_I, the corresponding $\Delta\mathcal{L}$ defines

[4] This expression follows from the definition of J and the fact that $\mathcal{I} = I$ and $\sigma_I = \sigma_{\mathcal{I}}$.

the *loudness just-noticeable difference* $\mathsf{JND}_{\mathcal{L}}$. Just as ΔI is commonly used to describe JND_I, we shall use $\Delta\mathcal{L}$ to describe $\mathsf{JND}_{\mathcal{L}}$. When the symbol $\Delta\mathcal{L}$ is used as an arbitrary loudness increment, rather than $\mathsf{JND}_{\mathcal{L}}$, one must carefully flag this unusual terminology.

While it is not possible to measure $\Delta\mathcal{L}$ (i.e., $\mathsf{JND}_{\mathcal{L}}$) directly, we assume that we may expand the loudness function in a Taylor series,[5] giving

$$\mathcal{L}(I+\Delta I) = \mathcal{L}(I) + \Delta I \frac{d\mathcal{L}}{dI}\bigg|_I + \mathsf{HOT},$$

where HOT represents *higher order terms* that we shall ignore (Allen and Neely 1997). If we solve for $\Delta\mathcal{L}$, defined as

$$\Delta\mathcal{L} \equiv \mathcal{L}(I+\Delta I) - \mathcal{L}(I), \tag{15}$$

we find

$$\Delta\mathcal{L} = \Delta I \frac{d\mathcal{L}}{dI}\bigg|_I. \tag{16}$$

We call this expression the *small-intensity-increment* approximation. It shows that the loudness JND $\Delta\mathcal{L}(I)$ is related to the intensity JND $\Delta I(I)$ by the slope of the loudness function evaluated at intensity I.

From the Taylor expansion the internal loudness standard deviation may be related to an external effective intensity variance by

$$\sigma_{\mathcal{L}} = \sigma_I \left|\frac{d\mathcal{L}}{dI}\right|.$$

It follows that $d' = \Delta\mathcal{L}/\sigma_{\mathcal{L}}$ and that $\mathsf{JND}_{\mathcal{L}}$ is defined by $d' = 1$ in a manner identical to the definition of the JND_I.

The loudness SNR. We define the *loudness SNR* as $\mathsf{SNR}_{\mathcal{L}} \equiv \mathcal{L}/\sigma_{\mathcal{L}}$. From the definitions of $\mathsf{SNR}_{\mathcal{L}}$, d', and J,

$$\mathsf{SNR}_{\mathcal{L}} = d' \left(J(I)\frac{d\mathcal{L}_{\log}}{d\beta}\right)^{-1}, \tag{17}$$

where $\beta \equiv 10\log_{10}(I/I_{\mathsf{ref}})$ and $\mathcal{L}_{\log}(\beta) \equiv 10\log_{10}(\mathcal{L}(10^{\beta/10}))$. This equation is important because *(a)* all the terms are dimensionless, *(b)* we are used to thinking of the loudness and intensity on a log scale (as in Fig. 1), and *(c)* $d\mathcal{L}_{\log}/d\beta$ is constant at large intensities (because, according to Stevens's law, $\mathcal{L}(I)$ is a power-law). To estimate the power-law slope using $\gamma \equiv d\mathcal{L}_{\log}/d\beta$ it is necessary to treat $\mathcal{L}$ as an intensity when defining $\mathcal{L}_{\log}$.

A much simpler way to write Eq. 17 is to define $\mathsf{SNR}_I \equiv I/\sigma_I$, which along with $\Delta I = d'\sigma_I$ gives

$$\mathsf{SNR}_I = \gamma \mathsf{SNR}_{\mathcal{L}}. \tag{18}$$

This equation says that the loudness SNR and the intensity SNR are related by the exponent γ of the loudness power-law function (**?**).

[5] While it is not meaningful to form a Taylor series expansion of the single-trial loudness $\widetilde{\mathcal{L}}(t, I)$, it is meaningful to expand the expected value of this random process.

2.3 Masking

Masking is the elevation in threshold due to a masking signal. To define masking we must first define the *masked threshold*. The energy of the masker spreads out along the basilar membrane with a density given by $I_x(f_m, I_m, x)$, where $x(f)$ is the *characteristic place* corresponding to frequency f. To model the masked threshold we need a model of $I_x(f_m, I_m, x)$ near the probe place $x(f_p)$.

The masked threshold. The hearing threshold in the presents of a masking signal is called the masked threshold. Since it is used in the definition of masking, it is a more fundamental than masking. More formally, the *masked threshold* $I_p^*(f_p, I_m)$ is the threshold intensity of a probe (maskee) I_p^* at frequency f_p in the presence of a masking signal having intensity I_m. When the masker intensity is set equal to zero, the masked threshold is just the probe intensity at the threshold of hearing in quiet, or the *unmasked threshold* $I_p^*(f_p) \equiv I_p^*(f_p, I_m = 0)$. As before, the asterisk indicates that special value of I_p which gives a 75% correct score for the detection of the probe in a 2-IFC task, due to the loudness uncertainty characterized by $\sigma_{\mathcal{L}}$.

Because the *hearing threshold* is generally defined statistically as the probe intensity corresponding to the 50% correct score (corrected for chance) for detecting the probe from some Ψ decision random variable, the masked threshold is *not* a random variable. To model masking we must first identify the Ψ decision random variable and then model the masked threshold $I_p^*(f_p, I_m)$ using the Theory of Signal Detection.

Masking and the masking pattern. The *masking* M is defined as the ratio of the masked to the unmasked threshold:

$$M(f_p, I_m) \equiv \frac{I_p^*(f_p, I_m)}{I_p^*(f_p)}.$$

The masked threshold is frequently reported in dB-SL (i.e., 10 $\log(M)$), where SL means that the masked threshold is referred to the *sensation level* (i.e., the unmasked threshold). The *masking pattern* is a description of the masking as a function of the masker level and the probe frequency. The masker can be any signal, such as a tone, narrowband noise, wideband noise, or even speech.

The masked threshold $I_p^*(f_p, I_m)$ is frequently measured with a pure-tone probe signal; however, a narrow probe band of noise centered on frequency f_p is sometimes used to reduce the beating that can take place when the masker is a pure tone. In this case it seems logical to measure the unmasked threshold $I_p^*(f_p)$ with the same probe signal when computing the masking.

2.3.1 Definition of $\Delta\mathcal{I}$ for masking.

We repeat the derivation of the intensity detector developed for $\text{JND}_{\mathcal{I}}$, but this time using a probe that differs from the masker. As in the derivation of Eq. 8, an increment in the intensity detector output $\Delta\mathcal{I}(\alpha)$ is defined in terms of a pressure scale factor α applied to the probe signal $\tilde{p}(t)$:

$$\widetilde{\mathcal{I}}(t, \alpha) \;=\; \widetilde{\mathcal{I}}(t, 0) + \widetilde{\Delta\mathcal{I}}(t, \alpha) \tag{19}$$

$$= \frac{1}{\rho c T}\int_{t-T}^{t}[\tilde{n}(t)+\alpha\tilde{p}(t)]^2 dt + \tilde{\epsilon}, \tag{20}$$

where $\tilde{n}(t)$ is the masker and $\tilde{p}(t)$ is the tone-probe (maskee). As a natural generalization of the Eq. 8, we set the intensity of the probe equal to that of the masker (i.e., $I = \sigma_p^2 = \sigma_n^2$), and control the intensity of the probe with the scale factor α. Expanding Eq. 20 and taking the expected value gives

$$\frac{\Delta\mathcal{I}(t,\alpha)}{\mathcal{I}} = \frac{2\alpha}{\rho c T}\int_{t-T}^{t}\rho_{np}(t)dt + \alpha^2, \tag{21}$$

where

$$\rho_{np}(t) \equiv \mathcal{E}[\tilde{n}(t)\tilde{p}(t)]/I \tag{22}$$

is the normalized correlation coefficient between the masker and probe. When ρ_{np} is nonstationary, it is a function of time t, and when it is stationary, it is constant over time, and can come out of the integral, which then integrates to 1. To simplify the notation, we define the effective correlation $\rho_e(t)$ as the integral of $\rho_{np}(t)$ over the T-second rectangular window,

$$\rho_e(t) \equiv \frac{1}{\rho c T}\int_{t-T}^{t}\rho_{np}(t)dt. \tag{23}$$

Equation 21 defines the relative size of the intensity detector's output $\Delta\mathcal{I}/\mathcal{I}$ as a function of α. If we require that $\Delta\mathcal{I}$ be at the detection threshold relative to the magnitude of the detector's internal noise $\tilde{\epsilon}$, then we may solve for α^*. In terms of the *de facto* detection measure d' [Eq. 11],

$$\frac{d'(\alpha)}{\mathsf{SNR}_I} \equiv 2\alpha\rho_e(t) + \alpha^2$$

Since $d'(\alpha^*) = 1$ defines α^*,

$$2\alpha^*\rho_e(t) + (\alpha^*)^2 = 1/\mathsf{SNR}_{\mathcal{I}}. \tag{24}$$

Because $\alpha \geq 0$ by definition, the solution to this equation is $\alpha^* = \sqrt{\rho_e^2(t) + 1/\mathsf{SNR}_{\mathcal{I}}} - \rho_e(t)$. The correlation between $\tilde{n}(t)$ and $\tilde{p}(t)$ is bounded between $-1 \leq \rho_{np}(t) \leq 1$, thus

$$\sqrt{\rho_e^2 + 1/\mathsf{SNR}_{\mathcal{I}}} - |\rho_e| \leq \alpha^* \leq \sqrt{\rho_e^2 + 1/\mathsf{SNR}_{\mathcal{I}}} + |\rho_e|.$$

This inequality bounds the range of $\alpha^*(\rho_e, \mathsf{SNR}_{\mathcal{I}})$ for the energy detector, for the case of $d' = 1$.

2.3.2 Classes of masking

The most basic classes of masking are simultaneous and nonsimultaneous masking. In this article we only consider simultaneous masking.

Frozen versus random maskers. Noise maskers come in two important forms: so-called frozen-noise and random-noise maskers. The term *frozen noise* is an oxymoron because the word *noise* is synonymous with stochastic, whereas *frozen* is synonymous with deterministic. We shall call such signals *high-degree-of-freedom signals*, or simply *frozen signals*, but never "frozen noise." Live music is an example of a stochastic signal, whereas recorded music is an

example of a high-degree-of-freedom signal.

As described by Eq. 14, the variance of a random masker can increase the masking. This effect has been called *external noise*. It is important to determine the relative contribution of the variance of the stimulus and the internal representation. This may be done by measuring the masked threshold twice, once with the random masker, and again with it frozen. If the two masked thresholds are the same, the internal noise is greater than the external noise.

Wideband maskers. When the masking signal has a widebandwidth the energy is spread out along the basilar membrane. For wideband signals, the degree of this correlation across frequency can be important in reducing the external noise. Because of the filtering and the nonlinear properties of the cochlea, it is necessary to understand narrowband masking before we attempt to analyze wideband maskers.

Narrowband maskers. When the masking signal has a narrow bandwidth, the spread of the energy along the basilar membrane is limited by the filtering properties of the cochlea. When the signal is deterministic or of long duration, the JND is limited by the internal noise.

There are two basic classes of narrowband masking measurements, called masking patterns (MPs), and psychophysical tuning curves (PTCs). The *masking pattern* is specified in terms of a fixed masked and a variable probe, while the *psychophysical tuning curve* is specified in terms of a fixed low-level (i.e., near-threshold) probe and a variable masker. Because of the nonlinear compressive properties of the cochlea, the difference between the MP and PTC, which is quite large, is important, as it gives insight into the nonlinear properties of the cochlea. We shall only deal with the MP here.

There are three basic regions of a masking pattern, corresponding to the *downward spread of masking* ($f_p < f_m$), the *upward spread of masking* ($f_p > f_m$), and *critical-band masking* ($f_p \approx f_m$). Critical-band masking is the realm of several poorly understood, but important, masking issues including the *linearity of masking* (an extension of Weber's law to the case of masking), the *asymmetry of masking* (the dB difference in masking between a tone and a narrow band of noise of equal intensity) and *beats*. When the frequency difference between the masker and the probe (maskee) becomes greater than the cochlear filter bandwidth, the masking depends on the shape of the cochlear filters and the cochlear nonlinear compression, which determine the properties of the upward and downward spread of masking.

Critical band masking and beats. Beating occurs when the masker and probe signals are correlated, as when two or more tones are within the bandwidth of a single cochlear filter (e.g., critical-band masking). This was the case for Riesz's 1928 experiment where the probe and masker were tones separated by 0.2 – 35 Hz. The presence of beats is quantified for the energy detector by $\rho_e(t)$. Within the cochlear filter bandwidth (i.e., the critical band) the signal pressure components add in a linear manner. It is frequently said that the power of the components adds, but this is incorrect; power adds only when $\rho_e = 0$, namely when there are no beats. Even though beat detection only occurs in a small frequency region around the critical band where the signal envelopes are correlated, it is critical to understand it quantitatively.

As the tones are moved apart in frequency, the signal develops a maximum roughness quality when the cochlear filter bandwidth is reached. This shows that the temporal integration

has a bandwidth that is greater than the critical bandwidth. When the frequency difference is greater than a critical band, the signals become independent ($\rho_e = 0$), the beating disappears, and the loudnesses of the masker and probe, presented together, add in magnitude, resulting in a total loudness that is always greater than the loudness of the masker alone.

Modulation detection. As may be seen from Eq. 21, $\Delta\mathcal{I}$ depends on two terms, a correlation term $\alpha\rho_e(t)$, and a fixed term α^2. When $\max|\rho_e(t)| > \alpha$, the correlation term dominates, and when $\max|\rho_e(t)| < \alpha$, the term quadratic term α^2 dominates. Thus when $f_m \approx f_p$, ΔI is time-varying around zero, and the *de facto* formula (Eq. 11) with $d' = 1$, derived under the assumption that $\Delta I > 0$, fails. When the mean loudness does not change (the case of critical band masking), the central nervous system must use a different criterion, which is characterized by the ratio of the variances, as described by a *maximum likelihood* analysis (Van Trees 1968). This critical-band detection paradigm is called *intensity modulation detection* (MD_I) to reflect the idea that the mean intensity is zero (or close to zero). Riesz's experiment, which is a special case of narrowband masking signals, is a classic example of MD_I (Viemeister 1988).

The masked threshold as an internal noise. In 1947 Miller pointed out that Riesz's "JND" experiment is formally a masking task, since the probe and the masker are not the same signal. He then demonstrated a close similarity between Riesz's modulation detection threshold and the wideband-noise JND_I. But practically speaking, Riesz's experiment is measuring a form of JND, because the masker and maskee are close in frequency (viz., 3 Hz apart). The term "close in frequency" is not well defined, but is related to the region of beats. The maximum rate of loudness variation is believed to be limited by an internal *temporal integrator* having a time constant between 100 and 300 ms, corresponding to a lowpass filter having a 3 dB bandwidth between 0.5 and 1.6 Hz (Yost 1994, section on "Temporal Integration"). For example, Munson (1947) found a single-pole integrator with an integration time constant of 200 ms, and Riesz showed that 3 Hz is the optimum frequency of modulation for his detection task. However, the 3 dB bandwidth is not a meaningful characterization of the perceptual bandwidth, since experiments show we can hear beats up to at least 20 or 30 Hz. This implies that the integrator should have a shallow slope and that perhaps the perceptual bandwidth should be specified at the -30 dB point rather that the -3 dB point, assuming Munson's single-pole filter. In summary, Riesz and Miller provided us with the insight that the masked threshold may be modeled in terms of the same (internal) noise that limits the JND (Littler 1965; Allen and Neely 1997).

It follows that the loudness uncertainty $\sigma_{\mathcal{L}}$ for Miller's wideband noise is similar in magnitude to $\sigma_{\mathcal{L}}$ for pure tones as measured by Riesz. Following Miller's lead, we define the masked threshold in terms of the Ψ uncertainty (e.g., "loudness noise"). If $\widetilde{\mathcal{L}}_m \equiv \widetilde{\mathcal{L}}(I_m)$ is the single-trial loudness due to a masker at intensity I_m, and $\widetilde{\mathcal{L}}(I_m, I_p)$ is the single-trial loudness due to both the masker and a probe of intensity I_p, then $\widetilde{\Delta\mathcal{L}}(I_m, I_p, \rho_e(t)) \equiv \widetilde{\mathcal{L}}(I_m, I_p) - \widetilde{\mathcal{L}}_m$ defines a decision variable for the masked threshold. The effective correlation $\rho_e(t)$ is used to account for the correlations between probe and masker corresponding to the critical-band region and beats. The probe intensity at the masked threshold $I_p^*(f_p, I_m)$ is defined as that value of I_p such that the probability of detecting the probe is 50% correct, corrected for chance.

3 The loudness SNR

We have interpreted the pure-tone JND as a measure of the Φ noise. In this section we complete this interpretation by calculating the loudness SNR required to account for the pure-tone and wideband noise JND$_I$. In the following we directly compare the tonal loudness growth function $\mathcal{L}(I)$ of Fletcher and Munson (1933) measured by Munson in 1933 to the tonal intensity JNDs $\Delta I(I)$ from Riesz (1928). Both sets of experimental data were taken in the same laboratory within a few years of each other, and it is likely they used the same methods and the same equipment, given its cost. This will allow us to estimate the loudness JND $\Delta\mathcal{L}(\mathcal{L})$, and therefore the *loudness signal-to-noise ratio* (SNR$_\mathcal{L} \equiv \mathcal{L}/\sigma_\mathcal{L}$). JND data are quite sensitive to the experimental measurement conditions (Stevens and Davis 1938, pages 141-143). The Riesz (Riesz 1928) and Munson (Munson 1932) data are interesting because they are taken under conditions similar to the loudness data of Fletcher and Munson, which were continuous (1 second long) pure tones.

3.1 A direct estimate of JND$_\mathcal{L}$

In Fig. 2 we present a direct estimate of the loudness JND$_\mathcal{L}$, [$\Delta\mathcal{L}(\mathcal{L})$] computed from Eq. 16 at all 11 frequencies Fletcher and Munson used to measure the loudness. The procedure for doing this is described in (Allen and Neely 1997). Each of the four panels displays a different frequency range. As indicated in the figure legend, we have marked the point on the curve where the slope changes. For the 62 Hz data in the upper-left panel we see that $\Delta\mathcal{L}$ is constant for levels below about 50 dB SL. Over most of the frequency range, below 20 dB [$\mathcal{L}(I) < 100$ LU], we find $\Delta\mathcal{L} \propto \sqrt{\mathcal{L}}$. Between 20 and 60 dB [$100 < \mathcal{L}(I) < 3000$], we find $\Delta\mathcal{L} \propto \mathcal{L}^{1/3}$. Above 60 dB [$\mathcal{L}(I) > 3000$], we find $\Delta\mathcal{L} \propto L$. Thus the loudness and JND$_\mathcal{L}$ are proportional above 60 dB SL.

In Fig. 2 on the lower-left we also show $\Delta\mathcal{L}(\mathcal{L})$ for Miller's (1947) wideband-noise JND$_I$ data. Miller gives the loudness-level as well as the intensity JND measurement. We converted this loudness-level to loudness using Fletcher and Munson's (1933) reference curve (i.e., Fig. 1). The SNR$_\mathcal{L}$ for the tones and the wideband noise are almost identical, especially over the frequency region 0.25 – 8.0 kHz.

3.2 Determination of the loudness SNR

Given that $d' \equiv \Delta\mathcal{L}/\sigma_\mathcal{L}$ and SNR$_\mathcal{L} \equiv \mathcal{L}/\sigma_\mathcal{L}$, it follows that $\mathcal{L}/\Delta\mathcal{L} =$ SNR$_\mathcal{L}/d'$. From Fig. 3 for levels above 65, the SNR$_\mathcal{L}$ becomes constant. From Fig. 2, SNR$_\mathcal{L}$ increases by a factor of 2 when the loudness increases by a factor of 4, up to about 55 dB.

As an application of Eq. 18, we calculate SNR$_\mathcal{L}$ for Miller's wideband masking data. Miller found J=0.1, which gives SNR$_I = d'/J = 10$. As shown in Fig. **??** on the upper right, the power-law has a slope of γ = 1/4 above 40 dB SL. Thus SNR$_\mathcal{L} \approx 40$ and this estimate is in fair agreement with estimates for pure tones as shown in Fig. 3.

Summary of JND$_\mathcal{L}$ results. The pure-tone and wideband-noise JND results have been summarized in terms of SNR$_\mathcal{L}(\mathcal{L})$. These curves seem similar enough that they may be characterized by one curve, at least for coding purposes. Between threshold and 60 dB SL, $\sigma_\mathcal{L} \propto \mathcal{L}^p$

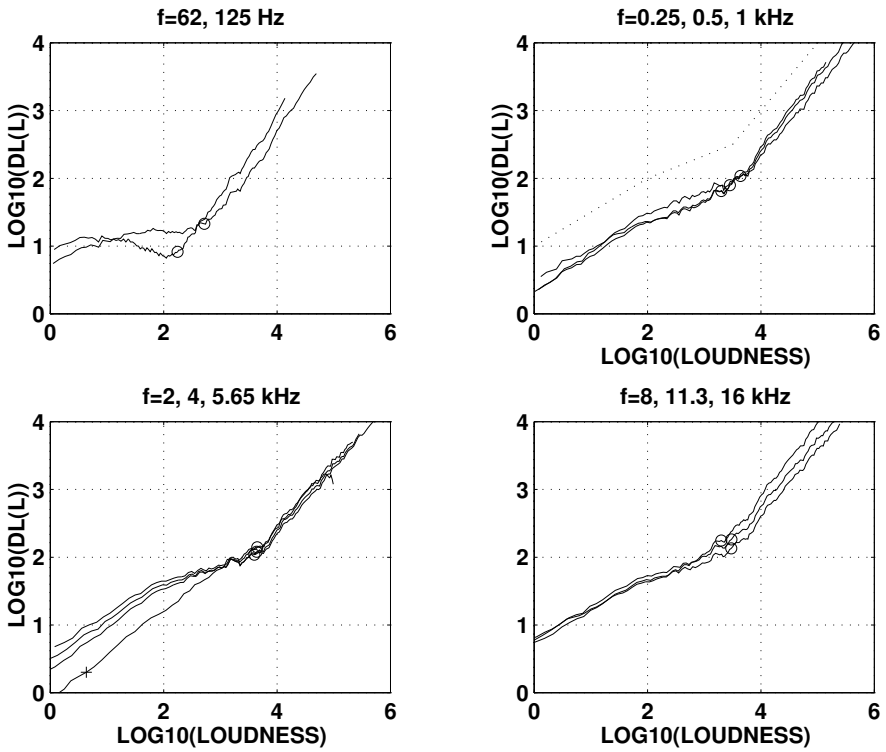

Figure 2: $\Delta\mathcal{L}(\mathcal{L}, f)$ computed directly from Eq. 16 using Riesz's JND data and the Fletcher-Munson loudness-intensity curve, for levels between 0 and 120 dB SL. The symbol $\odot$ has been placed on the curves at an intensity of 55 dB SL for 62 and 125 Hz, 60 dB SL for 0.25 to 1 kHz, 55 dB SL for 2 – 5.65 kHz, and 50 dB SL for 8 – 16 kHz. In the upper right panel we have added straight lines for reference, having slopes of 1/2, 1/3, and 1, for levels between 0 and 20 dB SL, 20 and 60 dB SL, and above 60 dB SL, respectively. From these plots it is clear that $\Delta\mathcal{L}(\mathcal{L})$ is described by a power law in $\mathcal{L}$ having three straight line segments. Between 0 and 20 dB SL, the slope is close to 0.5. Between 20 and 60 dB SL the slope is close to 1/3 ($\Delta\mathcal{L} \propto \mathcal{L}^{1/3}$). Above 60 dB SL, the slope is 1 ($\Delta\mathcal{L} \propto \mathcal{L}$). Fechner's hypothesis, that $\Delta\mathcal{L}$ is a constant [$\Delta\mathcal{L}(\not{I})$], appears to hold only for 62 and 125 Hz below 50 dB SL. One extra curve, labeled with the symbol +, has been added to the lower left panel, showing $\Delta\mathcal{L}(\mathcal{L})$ for the wideband noise case of Miller (1947). This curve has a slope of approximately 0.63 for $\mathcal{L}$ less than 10^3, and then merges with the tone data up to a loudness of 10^5, the upper limit of Miller's data.

with $1/3 \leq p \leq 1/2$ for tones and $p = 0.63$ for noise; above 60 dB SL, $\sigma_{\mathcal{L}} \propto \mathcal{L}$. Thus it appears that once we know the signal intensity, we know the loudness SNR for any signal bandwidth. Next we will look at masking data and describe how to use this information.

4 Narrowband maskers

Narrowband maskers, which result in a limited region of energy spread along the basilar membrane, hold the key to understanding masking. The two maskers we shall consider are a pure tone and a subcritical band of noise.

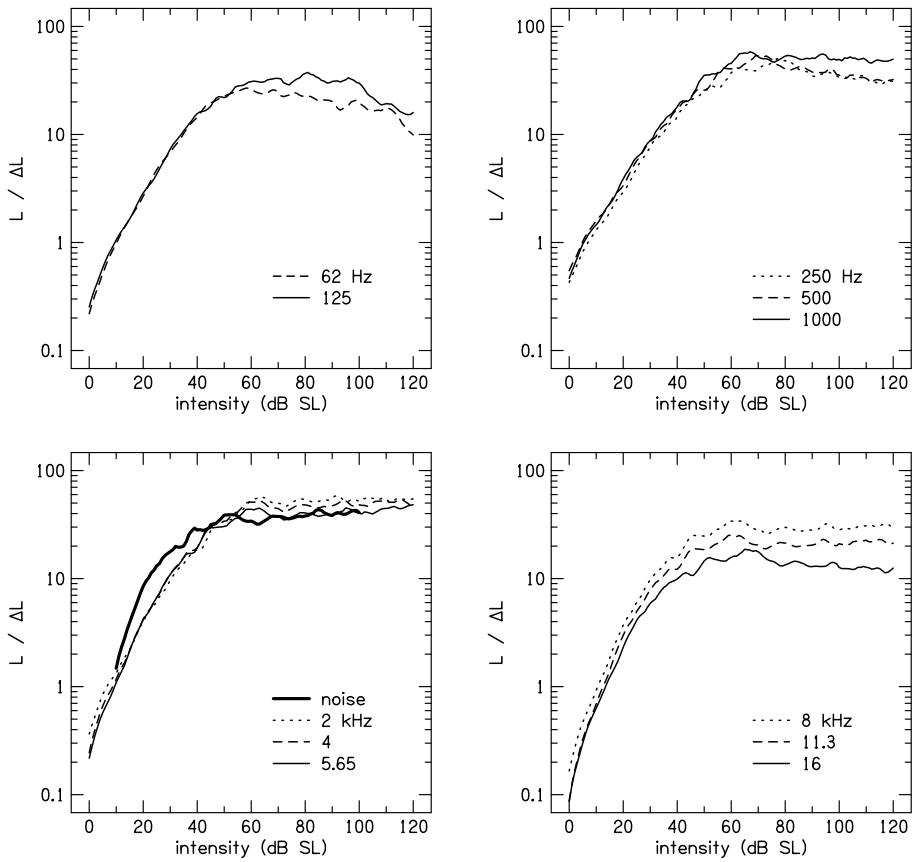

Figure 3: In this figure we plot $\mathcal{L}(I)/\Delta\mathcal{L}$ = SNR$_L/d'$ for intensities between 0 and 120 dB SL. Below about 55 dB SL the internal signal to noise ratio SNR$_{\mathcal{L}}(I)$ is increasing and is proportional to $\mathcal{L}^{1-1/p}$, where $2 \leq p \leq 3$ for tones and $p \approx 3/2$ for noise. Above 60 dB SL the SNR$_{\mathcal{L}}$ saturates at about 50 linear units. At 62 and 125 Hz the SNR$_{\mathcal{L}}$ decreases at high levels.

4.1 Tone-on-tone masking

When one tone is used to mask another tone, bandpass masking patterns result, as shown by the dashed curve of Fig. 5 (Egan and Hake 1950), corresponding to an $f_m = 400$ Hz tonal masker at 65 dB SL ($I_m = 80$ dB SPL), as a function of probe frequency f_p. Such patterns, as published by (Fletcher 1923a; Wegel and Lane 1924), were used by Fletcher and Munson to derive the theory of loudness (Fletcher and Munson 1933; Allen 1995). An alternative way to view these data is as *masking level curves* as shown in Fig. 6 for a f_m = 400 Hz masker. Such data were first published in (Fletcher 1923a; Fletcher 1923b; Wegel and Lane 1924) for frequencies f_p between 0.25 and 4 kHz and intensities between 0 and 85 dB SL. In 1922 Fletcher and Wegel published a major study that accurately measured the threshold of hearing for the first time (Allen 1995, Page A7), and the masked threshold measurements of Fig. 6 followed from this 1922 experiment.

It is interesting to compare the 1923 Fletcher data of Fig. 6 with the 1950 data of Fig. 5. For example, the masked threshold for a 65 dB SL masker is shown in Fig. 5 with **o** symbols. The largest difference is about 17 dB at 2 kHz. A careful comparison between the two data

sets shows that a 58 dB SL masker at 400 Hz from the 1923 Fletcher data (shown on Fig. 5 as **+** symbols) is within a few decibels of the 65 dB SL masker for Egan and Hake's data for all probe frequencies. Subject variability is one obvious explanation for these differences. An alternative is that Fletcher, Wegel, and Lane may have compensated for the $2f_1 - f_2$ distortion product in their measurements, as they did for the subjective harmonics.

The spread of masking. As may be seen from Fig. 6 (solid lines) for the case f_p = 2 – 4 kHz, the onset of masking is abrupt at about 60 – 65 dB and has a slope of about 2.4 as shown by the dotted line. The dotted superimposed on the 3 kHz curve is defined by:

$$\frac{I_p(f_p, I_m)}{I_{\text{ref}}} = \left(10^{-6}\frac{I_m}{I_{\text{ref}}}\right)^{2.4}.$$

This steep slope is referred to as the *upward spread of masking*. For *downward spread of masking* ($f_p < f_m$), the growth of masking is a compressive power law in intensity (dashed lines).

4.1.1 Critical-band masking

For probe frequencies near the masker frequency of 400 Hz the masking is said to be *linear in intensity*. For example, at f_p = 0.45 kHz (dash-dot line in Fig. 6) the masking curve is well approximated by the linear relation

$$\frac{I_p^*(f_p, I_m)}{I_m} = \frac{1}{40} \tag{25}$$

for I_m greater than about 25 dB SL, as indicated by the dotted line superimposed on the 0.45 Hz masking curve (Wegel and Lane 1924, page 270). Other examples of this linearity include the masking of tones by narrow bands of noise (Fletcher and Munson 1937; Fletcher 1938b; Egan and Hake 1950) and the masking of tones and narrow bands of noise by wideband noise (Fletcher and Munson 1937; Fletcher 1938a; Fletcher 1938b; Hawkins and Stevens 1950; Fletcher 1995).

Equation 25 is an extension of Weber's law for JNDs to the case of masking. It is just as important to understand (i.e., model) the linearity of masking as it is to understand why Weber's JND law holds for wideband noise, as the explanations are the same.

While the linearity of masking seems to be a trivial experimental observation, it is a surprising result. When the probe is added to the masker and the two signals are within a critical bandwidth, their basilar membrane motion adds (e.g., two sine waves beat). However, the response level of the basilar membrane motion, the neural response, and the resulting loudness are all nonlinear functions of level. Thus it is not initally obvious why the masking should be proportional to the intensity.

Linearity of masking and Weber's law. If J is approximately constant for $f_p = f_m$, then it is reasonable to expect that it will be approximately constant when $f_p \approx f_m$. If we interpret I_p as the change in intensity due to the probe then $\Delta I = I_p^*$ and $I_p^*/I_m \equiv \Delta I/I$. Thus Eq. 25

is an extension of Weber's JND law to masking, but is *not* Weber's law because that law strictly applies to the JND. Clearly however the two cases are functionally equivalent. Riesz was the first to recognized this important correspondence. Five years after Fletcher published the masking level curves, Riesz (Riesz 1928) executed an extensive quantitative study of the critical band region.

Riesz came to two important conclusions. *First*, he interpreted ΔI in terms of a short-term intensity variation, and defined $\Delta I = I_{\max} - I_{\min}$. With his interpretation of ΔI as a short-term intensity, he was able to precisely test Weber's law under conditions of masking. Thus Riesz's experiment was the first to make the important connection between critical-band masking and Weber's law.

Second, unlike Wegel and Lane's conclusion that I_p^*/I_m is a constant (i.e., that Weber's law holds), Riesz found that $\Delta I/I$ is not exactly constant. In other words, upon careful scrutiny, he showed that Eq. 25 does not hold exactly for the case of tones. Unfortunately it was almost 20 years before Riesz's observations were fully appreciated (Miller 1947; Littler 1965).

Riesz's JND experiment. According to Eq. 21, $\Delta\mathcal{I} = 0$ when the temporal integration time T is long relative to the time variations of $\rho_{np}(t)$ (Eq. 22). It is difficult to argue that ΔI is proportional to I (i.e., Weber's law) if $\Delta I = 0$.

Riesz found a trivial resolution of this problem. He assumed that the ear averages over a short enough interval that it can track the variations over time. This idea is obvious, because one can hear the slow beating of two sine waves as their loudness slowly varies. From this point of view, Riesz defined his measure of JND$_I$ as

$$\frac{\Delta I}{I} = \frac{(1+\alpha^*)^2 - (1-\alpha^*)^2}{(1-\alpha^*)^2},$$

which is $(I_{\max} - I_{\min})/I_{\min}$. For small α, Riesz's formula reduces to J $\approx 4\alpha$, which is similar to the first right-hand term in Eq. 21. When Riesz reports J = 0.1, we have α^* = 0.025.

If Riesz had ignored the beating and treated the two tones as independent, then ΔI would have been the intensity of the two tones played together minus the intensity of the masking tone alone, and he would have reported the Weber fraction as $J_i = [(1+\alpha^2) - 1]/1 = \alpha^2$, which is the second term in Eq. 21. Thus given his actual measure value of $\alpha^* = 0.025$, he would have reported J_i = 0.000625 rather than 0.1.

Intensity JND data have traditionally been expressed in many different ways, depending on the point of view of the author (Yost 1994, page 151 – 152). Because there have been so many different measures, there has been a great deal of confusion as to exactly what the numbers mean. The Weber fraction was originally defined to characterize the JND where the probe and masker are identical ($\rho_{np} = 1$). When applied to masking, J is a measure that depends on the effective correlation $\rho_e(t)$ and therefore on the temporal integration time. Until we determine how to precisely define the temporal integration time, it seems more appropriate to quote the experimental results in terms of α^* rather than J, because α^* does not depend on the independence assumption and therefore on $\rho_e(t)$.

Maximum likelihood formulation of Riesz's experiment. When two sine waves beat, Riesz's measure of $\Delta I = I_{\max} - I_{\min}$ is a reasonable statistic. However, we need a more general mea-

sure when dealing with arbitrary correlated critical-band signals. The method of maximum likelihood estimation is the natural way to do this (Van Trees 1968).

One could think of Riesz's experiment either in terms of a two-hypothesis test where H_0 is for $\alpha = 0$ and H_1 is for $\alpha > 0$, or as the detection of a 3 Hz sine wave in noise where $H_0 is \mathcal{N}(0, \sigma_{\mathcal{L}})$ and H_1 is the sine wave plus the same noise used in H_0. This is a modulation detection (MD_I) task where the means are equal and the variance changes. Thus when calculating the probability of a 75% correct response, we cannot use the *de facto* rule $d' = 1$, because this measure is always zero (because $\Delta\mathcal{L} = 0$).

When $\rho_e = 0$, a *sufficient statistic* is the ratio of the change in mean to the variance ratio (i.e., d'), while for the modulation detection case ($\rho_e \neq 0$), it is the variance ratio (Van Trees 1968). Given two normal distributions $\mathcal{N}(0, 1)$ and $\mathcal{N}(0, \sigma)$, the probability of correct classification by a maximum likelihood classifier is 0.742 when $\sigma = 3$. For the case of a sine wave in unit-variance noise [$\mathcal{N}(0, 1)$], simulations show that $\alpha^* \approx 6.0$. This then gives us a formal mechanism for relating the 2-IFC JND measurements to the modulation detection measurements.

4.2 Noise-on-tone maskers

The masking due to a tone and that due to a subcritical bandwidth of noise of equal intensity are very different. This difference has been called the *asymmetry of masking*. This asymmetry is clearly evident in Fig. 5 (Egan and Hake 1950, Figure 7) which compares a five-subject average masking pattern for a 90 Hz narrowband noise (solid curve) with the tone masking pattern. The intensity of both maskers is the same (80 dB SPL, or 65 dB SL). Fletcher (1995, page 205) showed that the loudness of a subcritical band of noise is the same as the loudness of a pure tone having the same intensity. Even though the intensity and the loudness are the same, from Fig. 5 the masked thresholds differ by about 20 dB at the masker frequency of 400 Hz, or by about 18 dB on either side (e.g., at 380 and 430 Hz).

Figures 3 and 4 in Egan and Hake's paper show single-subject results at 430 Hz where I_p^*/I_m = 1/1000 (i.e., -30 dB) for the tone masker, and 1/10 (i.e., -10 dB) for the noise masker, leading to a $10\log(10)$=20 dB difference off the masker frequency. Their figures 5 and 6 provide data for a second subject at three intensities as a function of frequency. At the masker frequency of 430 Hz the difference between the noise-masker and tone-masked threshold has a mean of 23 $\pm$ 1.6 dB.

We can use Eq. 24 to explain a significant portion of the asymmetry of masking. To do this we start with our estimate of $SNR_{\mathcal{L}}$ from Fig. 3. For a level of 80 dB SPL at 400 Hz, we have $SNR_{\mathcal{L}} \approx 40$. Since $\gamma \approx 1/3$ for tones, from Eq. 18 we estimate SNR_I to be about 13.3. From Eq. 24, with $|\rho_e| = 1$ (for tones), $\alpha^* = 1/26.6$, or -28.5 dB. The difference between the 65 dB SL tone masker level in Fig. 5 and the dashed line at 400 Hz is about -20 dB. The difference between -28.5 and -20 dB (-8.5 dB) represents the error in the prediction.

For the case of the noise masker, $\rho_e = 0$. From Eq. 24 $\alpha^* = 1/\sqrt{13.3}$ = 0.27, or -11.2 dB. The corresponding value from Fig. 5 is the difference between the noise masker level of 65 dB SL and the 61 dB level of the solid line near 410 Hz, which is -4 dB, resulting in a -7.2 dB error. The energy detector formula gives a 28.5-11.2 = 17.3 dB difference between the tone and narrowband-noise masker, compared to Egan and Hake's 20-4=16 dB difference. Thus while the absolute estimates of α^* are too small by about a factor of two (meaning either the estimate

of the loudness SNR may be too large), or the subjects were underperforming, the prediction of the asymmetry of masking is close to the measured value. There is some uncertainty in the value of the slope γ, since for noise it is 1/4, while for tones it is 1/3.

We conclude that the correlation between the masker and probe has a dramatic effect on the threshold signal gain α^*, with threshold intensity variations of up to 32.46 dB when $\text{SNR}_I \approx 10$ [i.e., from Eq. 24 with ρ_e between -1 and 1].

The energy detector analysis clearly show the importance of the correlation between the probe and masker. When a tone is added to a narrow band of noise of the same center frequency, the two signals move slowly in and out of phase, reflecting the correlation and increasing the variance of the decision variable. We conclude that a proper analysis of masking using a maximum likelihood analysis of the detection problem, applied in the loudness domain, will result in excellent correlations with masking experiments.

5 The loudness model

When a single tone is presented to the cochlea, the energy is spread out along the cochlea, even though the energy only exists at a single frequency. The function $H(f, x)$ defines a family of complex filter functions. Corresponding to every point x_0 there is a filter function $H(f, x_0)$, and for every pure tone at frequency f_0, the energy is spread along the basilar membrane according to $|H(f_0, x)|^2$. Assuming the signal is above the threshold at a given point x, the excitation at each point drives nerve fibers that innervate that patch of basilar membrane. The total spike rate for that patch defines the partial loudness rate $\mathcal{L}_{tx}(t, x)$. The total loudness is given by a double integral over time and place

$$\mathcal{L}(t) = \int_{\tau=-\infty}^{t} E_s(t-\tau) \int_{x=0}^{1} \mathcal{L}_{tx}(\tau, x) dx d\tau,$$

where $E_s(t)$ is Munson's (1947) sensation integral, and the integral over x is normalized to unit length. The partial loudness function $\mathcal{L}_{tx}(t, x)$ is a nonlinear transformation of the energy along the basilar membrane, defined by the bank of filters $H(f, x) = \mathcal{F} \cdot h(t, x)$ [$\mathcal{F}\cdot$ represents the Fourier transform, and $h(t, x)$ is a family of impulse responses].

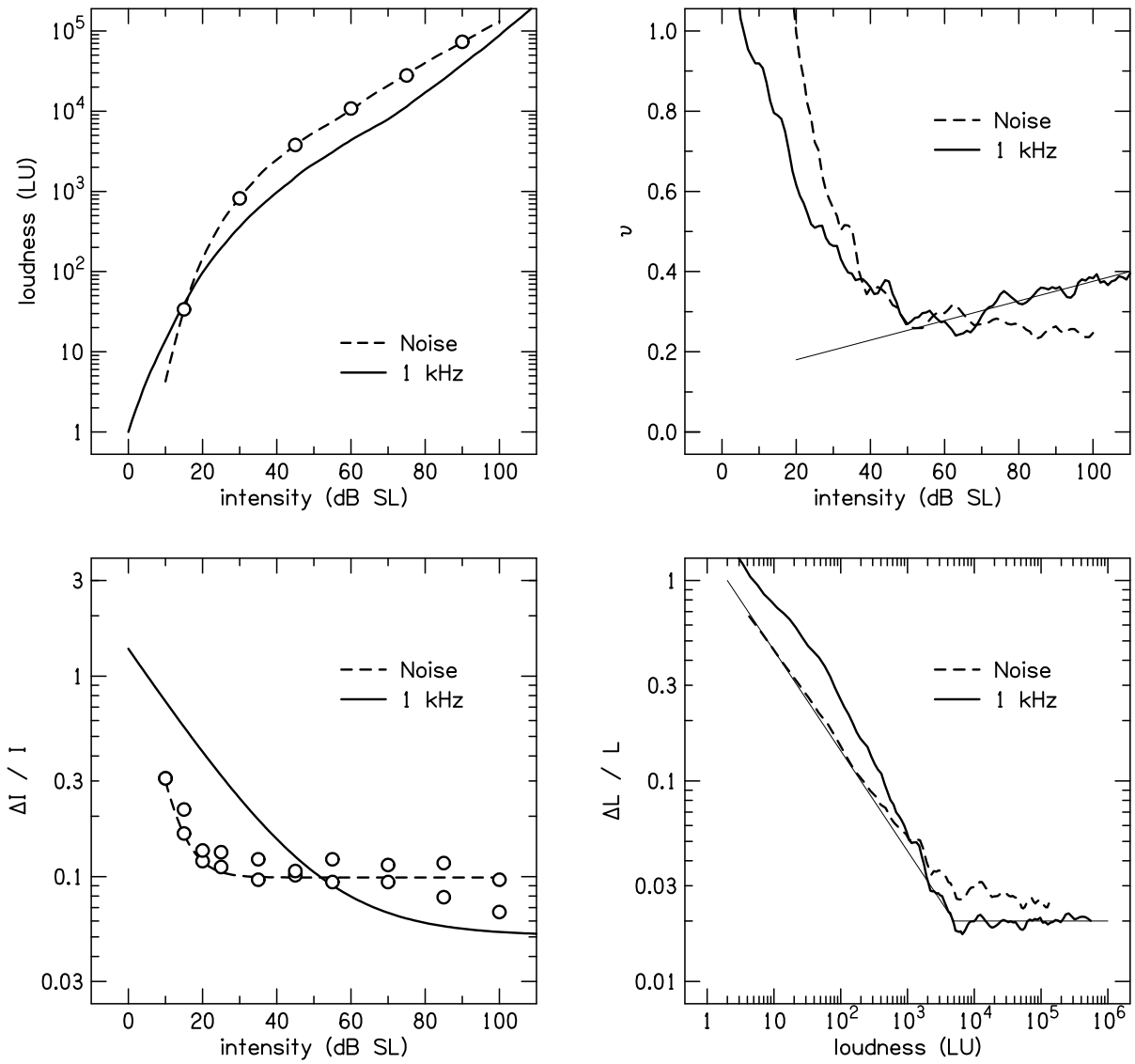

Figure 4: In 1947 Miller measured the JND_I and the loudness–level for two subjects using wideband noise (0.15–7 kHz) for levels between 3 and 100 dB SL. The intensity of the noise was modulated with a ramped square wave that was high for 1.5 s and low 4.5 s. The loudness, computed from Miller's phon data (dashed curve) using Fletcher and Munson's 1933 1 kHz tone loudness–growth curve are shown in the upper–left panel, along with the Fletcher Munson tonal loudness–growth function (solid curve). The upper–right panel shows the exponent $\gamma(I) \equiv d\mathcal{L}_{\log}/d\beta$ for both Fletcher and Munson's and Miller's (average of two subjects) loudness growth function. In the lower–left panel we plot $\Delta I/I$ versus I for Miller's two subjects, Miller's equation, and Riesz's equation. In the bottom–right panel we show the $\Delta\mathcal{L}/\mathcal{L}$ versus L for the noise and tones cases. From Eq. **??** $\Delta\mathcal{L}/\mathcal{L} = \gamma(I)J(I)$. Note how the product of $\gamma(I)$ and $J(I)$ is close to a constant for tones above 65 dB SL. This invariance justifies calling the variations in the power–law exponent $\gamma(I)$ for tones the "near–miss to Stevens' law." For reference, 1 sone is 975 LU.

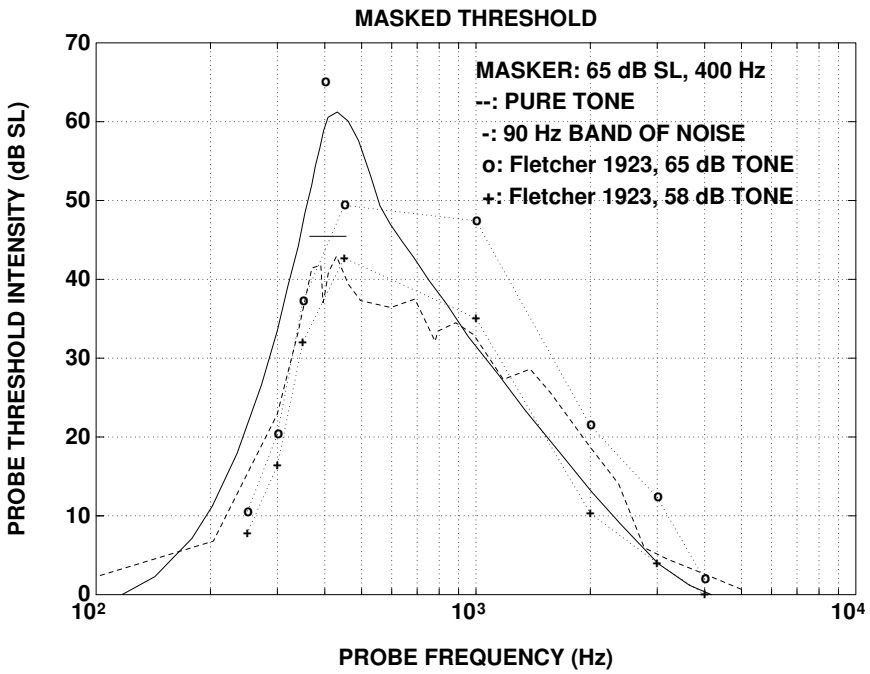

Figure 5: The solid curve is the simultaneous masking pattern for a 90 Hz band of noise centered at 410 Hz having an intensity of 65 dB SL (a spectral level of 45.6 dB, as shown by the short solid line). The dashed curve is the masking pattern for a 400 Hz pure tone at 65 dB SL (circle). The probe signal (maskee) was a 0.7 s pure tone. Note the large (26 dB) difference in the masked threshold at the masking frequency of 410 Hz. When the probe is near 500 Hz, the distortion product $2f_1 - f_2$ is the limiting factor in detection. The tone masking curve seems to be shifted to higher frequencies by a ratio of about 1.2 (a ratio of 600 and 500 Hz). The dips at 0.8 and 1.2 kHz are due to subjective harmonics (Fletcher 1995). The masked threshold for a 400 Hz tone as determined from the data of Fig. 6 of ref. (Fletcher 1923a) are shown by symbols **+** for 58 dB SL and **o** for 65 dB SL.

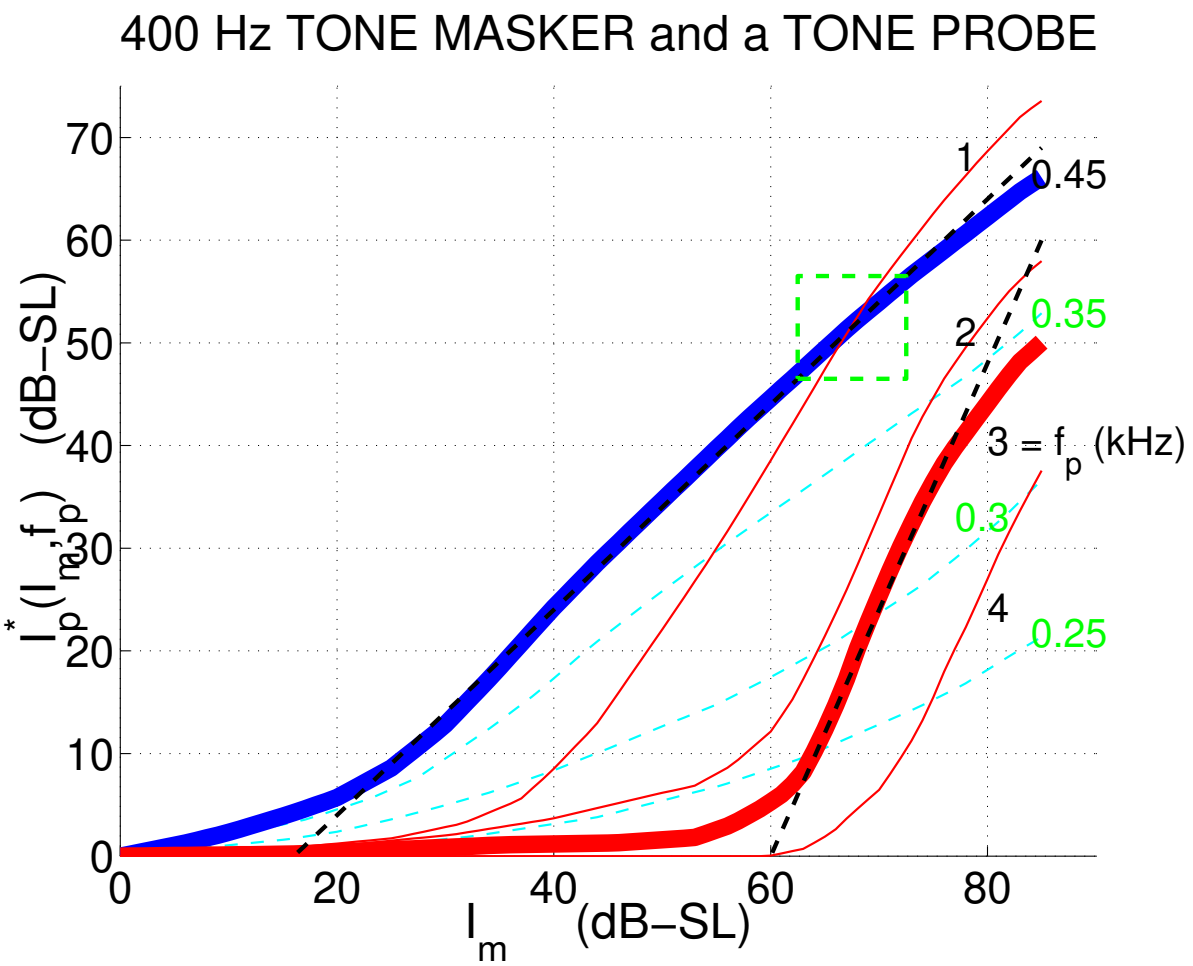

Figure 6: Tone-masking-tone data from Fletcher (1923) and Wegel and Lane (1924) for a masker at 400 Hz. The dashed lines correspond to probe frequencies between 0.25 – 0.45 kHz, while the solid lines correspond to probe frequencies of 1 – 4 kHz. The masking at 0.35 and 0.45 kHz is proportional to the masker level (i.e., the slope is close to 1). For 2, 3, and 4 kHz there is a threshold effect at about 60 dB SL. For these frequencies the slope is greater than 1. The short-dashed line superimposed on the 3 kHz curve is given by Eq. 4.1 and has a slope of 2.4 on a log-log axis. This steep slope is an important characteristic of the *upward spread of masking*.

Some basic questions. Here are some basic questions that are begging for further investigation:

- Why is the critical ratio independent of level (Fletcher 1938a; Fletcher 1995; Hawkins and Stevens 1950)?
- What is the relation between masking by narrow bands of frozen noise and the pure tone JND?
- Are Weber's law and the linear relation seen in narrowband and wideband maskers related?
- Why does the masked threshold track the masker intensity in a linear manner over such a large range of intensities, given that the BM at CF is nonlinear? That is, why is $\sigma_I \propto I$?
- Why do we find Weber's law to hold for wideband noise (Allen and Neely 1997)?
- What is the reason for the "near miss" to Weber's law for tonal signals (Yost 1994; Green 1988b; Allen and Neely 1997)?
- What is the source of the upward spread of masking (Wegel and Lane 1924; Ehmer 1959; Littler 1965)?
- Why is there a $>$ 26 dB difference in the masking between a tone and a narrow band of noise (Egan and Hake 1950)?
- What determines the ratio of the critical ratio to the ERB, and what is its true value?
- What are the relative contributions of internal and external noise in the masking of tones and narrow bands of noise, as a function of the bandwidth and duration of the signal?
- Why is there such a close relation between the frequency JND and the ERB (Fletcher 1995; Littler 1965)?
- Does loudness add (Marks 1979)?
- What is the loudness SNR for two equally loud tones that do not mask each other?

References

Allen, J. B. (1991). "Modeling the noise damaged cochlea," in Dallos, P., Geisler, C. D., Matthews, J. W., Ruggero, M. A., and Steele, C. R., editors, *The Mechanics and Biophysics of Hearing*, pages 324–332. Springer-Verlag, New York.

Allen, J. B. (1994). "How do humans process and recognize speech?," *IEEE Trans. on Speech and Audio Proc.* **2**(4):567–577.

Allen, J. B. (1995). "Harvey Fletcher 1884–1981," in Allen, J. B., editor, *The ASA edition of* Speech, Hearing in Communication, pages A1–A34. Acoustical Society of America, Woodbury, New York.

Allen, J. B. (1996). "DeRecruitment by multiband compression in hearing aids," in Jesteadt, W. and *et al.*, editors, *Modeling Sensorineural Hearing Loss*, pages 99–112. Lawrence Erlbaum, Inc., Hillsdale, NJ.

Allen, J. B. and Fahey, P. F. (1992). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," *Journal of the Acoustical Society of America* **92**(1):178–188.

Allen, J. B. and Neely, S. (1997). "Modeling the relation between the intensity JND and loudness for pure tones and wide–band noise," *Journal of the Acoustical Society of America* **102**(6):3628–3646.

Allen, J. B. and Neely, S. T. (1992). "Micromechanical models of the cochlea," *Physics Today* **45**(7):40–47.

Buus, S. (1990). "Level discrimination of frozen and random noise," *Journal of the Acoustical Society of America* **87**(6):2643–2654.

Carver, W. F. (1978). "Loudness balance procedures," in Katz, J., editor, *Handbook of clinical audiology, 2^{d} edition*, chapter 15, pages 164–178. Williams and Wilkins, Baltimore MD.

Egan, J. and Hake, H. (1950). "On the masking pattern of a simple auditory stimulus," *Journal of the Acoustical Society of America* **22**:662–630.

Ehmer, R. (1959). "Masking patterns of tone," *Journal of the Acoustical Society of America* **31**:1115–1120.

Fletcher, H. (1922). "The nature of speech and its interpretation," *Journal of the Franklin Institute* **193**(6):729–747.

Fletcher, H. (1923a). "Physical measurements of audition and their bearing on the theory of hearing," *Journal of the Franklin Institute* **196**(3):289–326.

Fletcher, H. (1923b). "Physical measurements of audition and their bearing on the theory of hearing," *Bell System Technical Journal* **ii**(4):145–180.

Fletcher, H. (1929). *Speech and Hearing*. D. Van Nostrand Company, Inc., New York.

Fletcher, H. (1938a). "Loudness, masking and their relation to the hearing process and the problem of noise measurement," *Journal of the Acoustical Society of America* **9**:275–293.

Fletcher, H. (1938b). "The mechanism of hearing as revealed through experiments on the masking effect of thermal noise," *Proceedings National Academy Science* **24**:265–274.

Fletcher, H. (1940). "Auditory patterns," *Reviews of Modern Physics* **12**:47–65.

Fletcher, H. (1995). "Speech and hearing in communication," in Allen, J. B., editor, *The ASA edition of Speech and Hearing in Communication*. Acoustical Society of America, New York.

Fletcher, H. and Galt, R. (1950). "Perception of speech and its relation to telephony," *Journal of the Acoustical Society of America* **22**:89–151.

Fletcher, H. and Munson, W. (1933). "Loudness, its definition, measurement, and calculation," *Journal of the Acoustical Society of America* **5**:82–108.

Fletcher, H. and Munson, W. (1937). "Relation between loudness and masking," *Journal of the Acoustical Society of America* **9**:1–10.

Fletcher, H. and Steinberg, J. (1924). "The dependence of the loudness of a complex sound upon the energy in the various frequency regions of the sound," *Physical Review* **24**(3):306–317.

Fletcher, H. and Wegel, R. (1922). "The frequency-sensitivity of normal ears," *Physical Review* **19**:553–565.

Gelfand, S. (1981). *Hearing, An introduction to psychological and physiological acoustics*. Marcel Dekker, Inc., New York and Basel.

Green, D. (1970). "Application of detection theory in psychophysics," *Proceedings of the IEEE* **58**(5):713–723.

Green, D. (1988a). "Audition: Psychophysics and perception," in Atkinson, R., Herrnstein, R., Lindzey, G., and Luce, R., editors, *Stevens' Handbook of Experimental Psychology*, chapter 6, pages 327–376. John Wiley & Sons, Inc., New York.

Green, D. (1988b). *Profile Analysis, Auditory Intensity Discrimination*. Oxford University Press, New York, Oxoford.

Hartmann, W. (1997). *Signals, Sound, and Sensation*. AIP Press, American Institute of Physics, Woodbury, NY.

Hawkins, J. and Stevens, S. (1950). "The masking of pure tones and of speech by white noise," *Journal of the Acoustical Society of America* **22**(1):6–13.

Hellman, W. and Hellman, R. (1990). "Intensity discrimination as the driving force for loudness. Application to pure tones in quiet," *Journal of the Acoustical Society of America* **87**(3):1255–1271.

Jayant, N. and Noll, P. (1984). *Digital coding of waveforms*. Prentice–Hall, Inc., Englewood Cliffs, NJ 07632.

Jesteadt, W., Wier, C., and Green, D. (1977). "Intensity discrimination as a function of frequency and sensation level," *Journal of the Acoustical Society of America* **61**(1):169–177.

Kingsbury, B. (1927). "A direct comparison of the loudness of pure tones," *Physical Review* **29**:588–600.

Knudsen, V. (1923). "The sensibility of the ear to small differences of intensity and frequency," *Phys. Rev.* **21**:84–103.

Liberman, M. and Kiang, N. (1978). "Acoustic trauma in cats," *Acta Otolaryngologica* **Suppl. 358**:1–63.

Liberman, M. C. and Dodds, L. (1984). "Single neuron labeling and chronic cochlear pathology III: Stereocilia damage and alterations of threshold tuning curves," *Hearing Research* **16**:55–74.

Littler, T. (1965). *The physics of the ear*. Pergamon Press, Oxford, England.

Lorente de No, R. (1937). "The diagnosis of diseases of the neural mechanism of hearing by the aid of sounds well above threshold," *Transactions of the American Otological Society* **27**:219–220. DISCUSSION OF E. P. FOWLER'S PAPER ON RECRUITMENT.

Marks, L. (1979). "A theory of loudness and loudness judgments," *Psycholgical Review* **86**(3):256–285.

McGill, W. J. and Goldberg, J. P. (1968). "Pure–tone intensity discrimination as energy detection," *Journal of the Acoustical Society of America* **44**:576–581.

Miller, G. A. (1947). "Sensitivity to changes in the intensity of white noise and its relation to masking and loudness," *Journal of the Acoustical Society of America* **19**:609–619.

Montgomery, H. C. (1935). "Influence of experimental technique on the measurement of differential intensty sensitivity of the ear," *Journal of the Acoustical Society of America* **7**:39–43.

Munson, W. (1947). "The growth of auditory sensation," *Journal of the Acoustical Society of America* **19**:584–591.

Munson, W. A. (1932). "An experimental determination of the equivalent loudness of pure tones," *Journal of the Acoustical Society of America* **4**(7):ABSTRACT.

Munson, W. A. and Gardner, M. B. (1950). "Loudness patterns – a new approach," *Journal of the Acoustical Society of America* **22**(2):177–190.

Neely, S. T. and Allen, J. B. (1996). "Relation between the rate of growth of loudness and the intensity DL," in Jesteadt, W. and *et al.*, editors, *Modeling Sensorineural Hearing Loss*, pages 213–222. Lawrence Erlbaum, Inc., Hillsdale, NJ.

Pickles, J. (1982). *An introduction to the physiology of hearing*. Academic Press Inc., London, England.

Riesz, R. (1928). "Differential intensity sensitivity of the ear for pure tones," *Physical Review* **31**(2):867–875.

Rosenblith, W. (1959). "Sensory preformance of organisms," *Reviews of modern physics* **31**:485–491.

Santos-Sacchi, J. and Dilger, J. P. (1987). "Whole cell currents and mechanical responses of isolated outer hair cells," *Hearing Research* **35**:143–150.

Siebert, W. (1965). "Some implications of the stochastic behavior of primary auditory neurons," *Kybernetik* **2**:205–215.

Steinberg, J. and Gardner, M. (1937). "Dependence of hearing impairment on sound intensity," *Journal of the Acoustical Society of America* **9**:11–23.

Stevens, S. and Davis, H. (1938). *Hearing, Its Psychology and Physiology*. Republished by the the Acoustical Society of America in 1983, Woodbury, New York.

Van Trees, H. (1968). *Detection, estimation, and modulation theory, Part 1*. John Wiley and Sons, New York.

Viemeister, N. F. (1988). "Psychophysical aspects of auditory intensity coding," in Edelman, G., Gall, W., and Cowan, W., editors, *Auditory Function*, chapter 7, pages 213–241. Wiley, New York.

Weber, E. H. (1988). "Der tastsinn und das gemainfül," in Wagner, R., editor, *Handwörterbuch der Physiologie*, volume 3, chapter 7, pages 481–588. Vieweg, Braunschweig.

Wegel, R. and Lane, C. (1924). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," *Physical Review* **23**:266–285.

Yost, W. (1994). *Fundamentals of Hearing, An Introduction*. Academic Press, San Diego, London.

Contents

1 Introduction **2**
1.1 The problem of perceptual coding . . . 2
1.2 Summary . . . 4
1.3 Overview of this article. . . . 5

2 Definitions **6**
2.1 Loudness . . . 6
2.1.1 Critical band . . . 7
2.1.2 Modeling loudness . . . 7
2.1.3 Loudness growth . . . 8
2.1.4 Loudness additivity . . . 9
2.2 The just-noticeable difference in intensity . . . 11
2.2.1 Definition of the JND_I . . . 13
2.2.2 Weber's law. . . . 14
2.2.3 Definition of $\Delta\mathcal{L}$. . . . 15
2.3 Masking . . . 17
2.3.1 Definition of $\Delta\mathcal{I}$ for masking. . . . 17
2.3.2 Classes of masking . . . 18

3 The loudness SNR **21**
3.1 A direct estimate of $\mathrm{JND}_{\mathcal{L}}$ . . . 21
3.2 Determination of the loudness SNR . . . 21

4 Narrowband maskers **22**
4.1 Tone-on-tone masking . . . 23
4.1.1 Critical-band masking . . . 24
4.2 Noise-on-tone maskers . . . 26

5 The loudness model **27**